

# MODELLKONTROLL FOR KASUS-KONTROLL STUDIER VED MARTINGAL RESIDUAL PROSESSER

av

TRYGVE RAGNAR SUNDE

**MASTEROPPGAVE**

*for graden*

***Master i Modellering og dataanalyse***

*(Master of Science)*



*Det matematisk- naturvitenskapelige fakultet  
Universitetet i Oslo*

*November 2008*

*Faculty of Mathematics and Natural Sciences  
University of Oslo*



# Forord

Denne oppgaven er skrevet som en del av mastergraden min i modellering og dataanalyse ved Universitetet i Oslo. Oppgaven er skrevet i perioden 2006 til 2008. Veileder for oppgaven har vært professor Ørnulf Borgan.

Først vil jeg takke Ørnulf Borgan som med smittende engasjement har hjulpet og støttet meg gjennom hele prosessen. Hans hjelpsomme kommentarer og positive innstilling har vært med på å gjøre det til en glede å jobbe med oppgaven i hele perioden.

Takk til Terje Kvernes ved matematisk institutt som hjalp meg med å få tilgang til tilstrekkelig mange datamaskiner til simuleringene, og for hjelp med tekniske problemer underveis.

Takk til alle venner og kjente som har hjulpet meg gjennom prosessen. Spesielt stor takk til alle på b800 for hyggelig selskap og god hjelp i disse to årene. Til slutt vil jeg takke familien min som alltid har vært der for meg.

Oslo, November 2008,  
*Trygve Ragnar Sunde*



# Innhold

<b>Forord</b>	<b>i</b>
<b>1 Innledning</b>	<b>1</b>
<b>2 Overlevelsesanalyse</b>	<b>5</b>
2.1 Overlevelsesdata . . . . .	5
2.1.1 Overlevelsesstudier og overlevelsesdata . . . . .	5
2.1.2 Begreper for overlevelsesdata . . . . .	6
2.1.3 Kovariater og regresjonsmodeller for overlevelsesdata . . . . .	7
2.1.4 Sensurering og trunkering . . . . .	7
2.1.5 Gjentatte hendelser . . . . .	9
2.1.6 Statistiske metoder for kohortdata . . . . .	9
2.2 Telleprosess formulering av kohort data . . . . .	15
2.2.1 Når modellen stemmer . . . . .	17
<b>3 Kasus-kontroll studier</b>	<b>19</b>
3.1 Innledning . . . . .	19
3.2 Sampling . . . . .	20
3.3 Nøstet kasus-kontroll design . . . . .	21
3.3.1 Enkel nøstet kasus-kontroll design . . . . .	21
3.3.2 Counter-matcing . . . . .	22

3.4	Statistiske metoder for kasus-kontroll . . . . .	23
3.5	Telleprosessformulering av kasus-kontroll . . . . .	26
3.6	Grupperte martingal residualprosesser . . . . .	29
3.7	Kji-kvadrat goodness-of-fit test . . . . .	32
<b>4</b>	<b>Om simuleringene</b>	<b>35</b>
4.1	Innledning . . . . .	35
4.2	Simulering av levetider . . . . .	35
4.2.1	Om sensurering i simuleringene . . . . .	37
4.2.2	Om tidsbruk i simuleringene . . . . .	41
4.3	Utvidelser av simuleringsalgoritmen . . . . .	41
4.3.1	Tidsavhengig effekt av kovariatene . . . . .	42
4.3.2	Feil koding av kovariat . . . . .	43
4.4	Problemer i simuleringene . . . . .	44
4.4.1	Negativ risiko . . . . .	45
4.4.2	Divergens i scoring-algoritmen . . . . .	47
<b>5</b>	<b>Simuleringsresultater</b>	<b>51</b>
5.1	Innledning . . . . .	51
5.2	Karakteristikker ved data . . . . .	51
5.3	Grupperinger i testen . . . . .	54
5.4	Når modellen stemmer . . . . .	56
5.4.1	Regresjonsanalyse av signifikansnivået . . . . .	58
5.5	Når modellen ikke stemmer . . . . .	63
5.5.1	Feil modell tilpasset. Liten forskjell på modellene. . . . .	64
5.5.2	Feil modell tilpasset. Stor forskjell på modellene. . . . .	66
5.5.3	Tidsavhengige effekt av kovariat. Liten positiv effekt. . . . .	70

5.5.4	Tidsavhengig effekt av kovariat. Stor positiv effekt. . . . .	72
5.5.5	Manglende kovariat. . . . .	76
5.5.6	Feil koding av kovariat. . . . .	78
<b>6</b>	<b>Diskusjon</b>	<b>81</b>
<b>A</b>	<b>Vedlegg</b>	<b>85</b>
	<b>Kilder</b>	<b>103</b>





# Kapittel 1

## Innledning

I overlevelsesanalysen studerer man hvordan ulike variabler påvirker levetiden til individer. Mer generelt studerer vi tid til en hendelse av interesse. Denne hendelsen trenger ikke å være død, men kan like gjerne være en kvinnes første fødsel. Overlevelsesanalysen har en rekke anvendelser i blant annet samfunnsfag, forsikring og medisin. I denne oppgaven vil vi konsentrere oss om de medisinske anvendelsene. Selv om resultatene vil være overførbare til andre områder vil vi studere situasjoner som er mest relevant i forhold til medisinske studier. I medisinske studier følges et antall individer over et gitt tidsrom. I løpet av studietiden observeres alle dødsfall i kohorten. Sammen med informasjon om individenes verdier av enkelte forklaringsvariable (kovariater), kan vi dermed studere sammenhengen mellom forklaringsvariablene og levetidene. Når en studie avsluttes, lever ofte mange av individene fortsatt. Om levetidene til disse individene vites bare at de minst varer til tidspunktet da studien ble avsluttet. Enkelte individer kan også forsvinne ut av studien av andre årsaker enn død. For eksempel kan de flytte eller pådra seg en sykdom som gjør det vanskelig å fortsatt være med i studien. Vi får dermed heller ikke anledning til å observere deres levetid. Individer som vi av ulike årsaker ikke får anledning til å observere dødstidspunktet til sier vi at er sensurert. På grunn av at sensurering i praksis alltid er tilstede i analyse av levetidsdata har vi behov for spesielle metoder som kan brukes på sensurerte data.

Hvis vi kun er interessert i effekten av en eller to kategoriske kovariater kan vi analysere deres effekt på overlevelsen til individene ved å gruppere individene etter disse kovariatene og studere forskjellene mellom gruppene. Vanligvis er vi interessert i flere enn to forklaringsvariabler og noen av disse kovariatene vil typisk være numeriske. For å trekke slutninger om kovariatenes effekt på overlevelsen i slike situasjoner må vi bruke regresjonsmodeller. Den mest brukte regresjonsmodellen i overlevelsesanalyse er Cox regresjonsmodell. Denne modellen antar at hasardraten til et individ med kovariater  $x_1, \dots, x_p$ , er gitt som

$$\alpha(t|x_1, \dots, x_p) = \alpha_0(t) \exp \{ \beta_1 x_1 + \dots + \beta_p x_p \},$$

der  $\alpha_0(t)$  er baseline hasard, mens  $\exp \{ \beta_1 x_1 + \dots + \beta_p x_p \}$  er relativ risiko, som beskriver

sammenhengen mellom kovariater og hasrardrate. For å tilpasse modellen til et datasett estimeres regresjonskoeffisientene  $\beta_1, \dots, \beta_p$  ved  $\hat{\beta}_1, \dots, \hat{\beta}_p$ . Disse estimatene for regresjonskoeffisientene finner vi ved å sammenlikne kovariatverdiene til individene som dør, med kovariatverdiene til individene som ikke er døde.

Cox regresjonsmodell baserer seg på at vi kjenner kovariatverdiene til alle individer i kohorten, såkalte kohort studier. Ved Cox regresjon sammenligner vi kovariatverdiene til de individene som dør (kasusene) med kovariatverdiene til alle individene som lever. Å fremskaffe den nødvendige informasjonen for alle individene i kohorten kan være vanskelig og svært kostbart. Dette gjelder særlig i epidemiologiske studier av sjeldne sykdommer. Her er det nødvendig å følge et stort antall individer slik at vi er sikre på at tilstrekkelig mange dødsfall vil inntreffe. Metoder som baserer seg på å sammenlikne kovariatverdiene til individene som dør med et mindre utvalg kontroller, kan derfor være godt alternativ. Slike studier kalles kasus-kontroll studier.

Alle modeller er tilnærmelser til virkeligheten, men noen modeller er bedre enn andre. Ved å velge gal modell kan vi risikere å misforstå sammenhengen mellom forklaringsvariable og overlevelse. Vi har derfor behov for verktøy som kan evaluere modellen som er valgt.

Det finnes ulike metoder for å evaluere tilpasningen av modeller til overlevelsedata. Både grafiske metoder og mer formelle tester er tilgjengelige. Disse er for eksempel beskrevet i Klein og Moschberger (2003). De fleste metoder er utviklet for kasus-kohort studier. Blant mulighetene er metoder basert på estimerte martingal residual prosesser, presentert av Barlow og Prentice (1988). Aalen (1993) viste hvordan plot av estimerte grupperte martingal residual prosesser mot tid kunne brukes som en sjekk av modelltilpasning for Aalens additive modell. Grønnesby og Borgan (1996) utvidet denne teorien til Cox-modellen og utviklet en formell goodness-of-fit test basert på estimerte grupperte martingal residual prosesser. For kasus-kontroll studier er metoder for modellkontroll mindre utviklet, men Borgan og Langholz (2007) viste hvordan goodness-of-fit testen til Grønnesby og Borgan kan utvides til kasus-kontroll studier.

Formålet med denne oppgaven er å studere egenskapene til goodness-of-fit testen til Grønnesby og Borgan for kasus-kontroll studier. Vi vil undersøke om fordelingsantakelsene under nullhypotesen holder, og vi vil undersøke teststyrken i ulike situasjoner. I studier av levetidsdata er det noen faktorer som kan tenkes å påvirke analysen vår. Andel sensurerte levetider og antall kontroller som samples er eksempler på slike faktorer. Vi vil også studere om signifikansnivå eller teststyrke påvirkes av noen slike faktorer som opptrer i virkelige kasus-kontroll studier. Siden goodness-of-fit testen baserer seg på estimerte grupperte martingal residual prosesser vil vi også undersøke om ulike grupperinger av de estimerte martingal residual prosessene gir ulike resultater. Ved siden av testen vil vi bruke plot av estimerte martingal residual prosesser til å evaluere modelltilpasningen.

Oppgaven er organisert på følgende måte. Vi vil begynne med å gi en kort innføring i overlevelseseanalyse i kapittel 2. Her vil vi presentere verktøy for å studere levetidsdata som baserer seg på kjennskap til kovariatverdiene til hele kohorten. I kapittel 3 vil vi vise hvordan disse analysemetodene kan anvendes på kasus-kontroll studier. To ulike design for sampling av kontrollene vil bli presentert. Det vil bli vist hvordan plot av estimerte grupperte martingal residual prosesser kan brukes for å evaluere hvor godt en modell passer

til data. Goodness-of-fit testen som vi skal studere egenskapene til i denne oppgaven, vil også bli introdusert her.

Ved å simulere levetider kan vi undersøke hvordan goodness-of-fit testen fungerer i ulike situasjoner. I simuleringene våre kan vi selv bestemme andel sensurerte levetider og antall kontroller for eksempel. Dette gir oss muligheten til å studere om egenskapene til testen påvirkes av slike faktorer. I kapittel 4 vil vi gjøre rede for hvordan levetider kan simuleres. Det vil også bli vist hvordan vi kan gi de simulerte levetidene ulike egenskaper som vi ønsker å studere i forhold til goodness-of-fit testen senere. Problemer som er oppstått i simuleringene er også tatt opp i dette kapitlet.

Kapittel 5 begynner med å ta for seg ulike faktorer vi ønsker å variere i simuleringene. Deretter blir det gjort rede for hvilke ulike grupperinger som vil bli brukt i goodness-of-fit testen. Vi vil fortsette med å gjøre rede for signifikansnivået til testen i avsnitt 5.4. Dette gjør vi ved å se på fordelingen til testobservatorene når vi tilpasser den sanne modellen til data. Nullhypotesen vi tester med goodness-of-fit testen er alltid at riktig modell er tilpasset. I avsnitt 5.5 ser vi på testens evne til å forkaste nullhypotesen når denne er feil. Dette er styrken til testen. Når vi simulerer kan vi gi data de egenskaper vi ønsker og vi vet hva som er den riktige modellen. Ved å tilpasse en modell vi vet at er feil kan vi undersøke styrken til testen i ulike situasjoner.

En oppsummering av simuleringsresultatene samt konklusjoner kommer i kapittel 6. Her vil det også bli foreslått mulige fortsettelser av studiene våre. Vedlegget inkluderer resultater fra simuleringene våre.



## Kapittel 2

# Overlevelsesanalyse

I dette kapittelet vil vi presentere verktøy for å analysere overlevelsesdata. Metodene som introduseres her forutsetter at vi kjenner kovariatverdiene til alle individene i kohorten. I kap 3 vil vi utvide denne teorien til også å gjelde kasus-kontroll studier. Stoffet i dette kapittelet er hentet fra Aalen, Borgan og Gjessing (2008).

### 2.1 Overlevelsesdata

#### 2.1.1 Overlevelsesstudier og overlevelsesdata

I overlevelsesanalysen studerer vi tiden til død hos et individ eller en kohort. En kohort er en gruppe individer som deler en eller flere egenskaper, som for eksempel at de har hatt kreft. Motivasjonene for slike studier kan være flere, men vi vil hovedsaklig konsentrere oss om epidemiologiske anvedelser. Dette er for eksempel medisinske undersøkelser der vi studerer befolkningsdata for å avdekke risikofaktorer eller årsakssammenhenger. Et nyere eksempel er analyse av levetidene til individer som har lagret biologisk materiale i biobanker.

Ord som overlevelsesdata og levetider kan virke misvisende. Disse trenger ikke henvise til det vi vanligvis tenker på som levetider, det vil si tid fram til død. Derimot kan det være tiden fram til en kvinne føder sitt første barn eller tid fra giftemål til skilsmisse. Likevel er det vanlig å omtale disse tidene som levetider.

Vi skiller mellom kalendertid og undersøkelsestid eller studietid. Undersøkelsestiden kan begynne å løpe for to personer ved ulike kalendertid. Ofte begynner undersøkelsestiden å løpe når en person blir med i undersøkelsen, men andre ganger kan man velge å la undersøkelsestiden begynne ved fødselen, selv om man kommer inn i undersøkelsen senere.

Overlevelsesdata kan vi samle inn i forbindelse med medisinske studier der vi for eksempel medisinerer halvparten av forsøkspersonene med et kolesterolsenkende middel, mens resten

får placebo, og deretter følges de to gruppene opp i en tid etterpå. Overlevelsesdataene kan også hentes i offentlige arkiver slik som kreftregisteret.

### 2.1.2 Begreper for overlevelsesdata

Overlevelsesfunksjonen gir sannsynligheten for at et individ enda ikke har opplevd hendelsen vi studerer ved tid  $t$  i studietid. Overlevelsesfunksjonen blir da

$$S(t) = P(T > t),$$

der  $T$  er individets levetid. Vi har at

$$S(t) = 1 - F(t)$$

der  $F(t) = P(T \leq t)$  er den kumulative fordelingsfunksjonen. Sannsynlighetstettheten  $f(t) = F'(t)$  multiplisert med  $dt$  gir dødssannsynligheten i det korte intervallet  $[t, t + dt)$  for et individ i gruppen,

$$f(t)dt \approx P(t \leq T < t + dt)$$

Hasardraten til et individ ved tid  $t$  kaller vi  $\alpha(t)$ . Denne er definert som:

$$\alpha(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T < t + dt | T \geq t)$$

Merk at  $\alpha(t)dt$  er den betingete sannsynlighetene for at et individ som fortsatt lever ved tid  $t$ , vil dø i det neste korte tidsintervall  $[t, t + dt)$ . Ved å integrere hasardfunksjonen finner vi den kumulative hasarden.

$$A(t) = \int_0^t \alpha(u)du$$

Vi har følgende sammenheng mellom hasardraten og overlevelsesfunksjonen:

$$\alpha(t) = -\frac{f(t)}{S(t)} = \frac{d}{dt} \log S(t). \quad (2.1)$$

Overlevelsesfunksjonen fås ved å integrere (2.1)

$$S(t) = e^{-A(t)} = e^{-\int_0^t \alpha(u)du}. \quad (2.2)$$

Merk at vi kan ha at  $S(\infty) > 0$ , siden et individ ikke nødvendigvis opplever den hendelsen vi er interessert i. Ikke alle får for eksempel kreft. I såfall er  $T$  en degenerert stokastisk variabel.

### 2.1.3 Kovariater og regresjonsmodeller for overlevelsesdata

Vi ønsker ofte å studere hvilken effekt en eller flere kovariater eller forklaringsvariabler har på hvordan hendelser inntreffer. I epidemiologiske undersøkelser er vi ofte ute etter risikofaktorer for eksempel for kreft eller influensa. Kovariatene kan være alt som kan måles, slik som alder, luftforurensning, krefttykkelse, kjønn og kolestrolnivå. Når vi kjenner til sammenhengene mellom risikofaktorer og sykdommer blir det lettere å unngå disse risikofaktorene i fremtiden. Et klassisk eksempel er sammenhengen mellom tobakk og kreft.

Om kovariatene antar vi at de er forutsigbare. Det betyr at enten er de faste og kan da måles i starten av studien, eller så er de tidsavhengige, og vi antar da at verdien til en kovariat ved tid  $t$  er kjent rett før tid  $t$ .

Vi skiller mellom interne og eksterne kovariater. Eksterne kovariater er enten faste, som for eksempel alder målt ved starten av studien, men de kan også være tidsavhengige. Eksterne tidsavhengige kovariater varierer uavhengig av om hendelsen vi interesserer oss for inntreffer eller ikke. Et eksempel på en slik kovariat er luftforurensning. Tidsavhengige kovariater som ikke er eksterne er interne. Dette er gjerne målinger vi utfører på forsøkspersonene i løpet av studien, slik som blodtrykk.

**Eksempel 1 (Kovariater)** Mellom 1962 og 1969 ble 488 pasienter med skrumplever ved flere sykehus i København inkludert i et randomisert forsøk. Formålet med studien var å undersøke om pasienter behandlet med hormonet prednison, hadde bedre overlevelse enn pasienter som fikk en uvirksom placebobehandling. 251 av pasientene fikk prednison, mens 237 fikk placebo. Studien pågikk fram til 1. oktober 1974. Tabell 1 gir en oversikt over kovariatene. Alder og protrombin index er delt i inn i grupper. Protrombin index er et mål på leverfunksjon. Ascites er væskeoppsamling i bukhulen.

I dette eksemplet har vi bare eksterne kovariater, nemlig behandling, alder ved starten av studien, kjønn, ascites ved behandlingsstart og protrombin indeks ved behandlingsstart. Alle kovariatene er faste og tidsuavhengige. Med dette mener vi at de ikke forandrer seg med tiden. Dette gjelder hva slags behandling pasientene fikk, hvilken alder de hadde ved behandlingsstart, deres kjønn, ascites ved behandlingsstart og protrombin indeks ved behandlingsstart. Dersom vi hadde supplert analysen med målinger av for eksempel protrombin index underveis i oppfølgingen hadde dette blitt en intern kovariat.

### 2.1.4 Sensurering og trunkering

Sensurering er noe vi omtrent alltid får når vi gjennomfører levetidsundersøkelser som beskrevet ovenfor. At en person blir sensurert vil si at vi slutter å observere personen før hendelsen vi studerer har inntruffet. Grunnene til sensurering kan være mange, og kommer an på hvilken undersøkelse vi gjennomfører. Mange medisinske studier forutsetter for eksempel at forsøkspersonene jevnlig kan testes hos en lege. Dersom en av forsøkspersonene flytter midt i en studie kan det bli vanskelig å holde kontakten. Vi må være sikre på at

Tabell 2.1: Kovariatene i skrumplevereksemplet.

Kovariat	Antall	Prosent
Behanding		
Placebo	237	49
Prednison	251	51
Kjønn		
Menn	290	59
Kvinner	198	41
Ascites		
Intet	386	79
Lite	54	11
Moderat eller mye	48	10
Alder		
under 50	91	19
50-59	163	33
60-69	178	36
over 70	56	11
Protrombin index		
under 50	116	24
50-69	154	32
70-89	122	25
over 90	96	20

vi får tak i den informasjonen vi er ute etter. Andre grunner kan være at man av en eller annen grunn ikke ønsker å delta i studien lenger eller at man blir rammet av annen sykdom som gjør det umulig å fortsette studien. Til slutt blir også alle som fortsatt ikke har opplevd hendelsen vi studerer når studien avsluttes sensurert, da vi følgelig ikke får observert deres levetider. Det eneste vi vet er at hendelsen vi studerte ikke inntraff før studien ble avsluttet. De ufullstendige levetidsdataene kalles de sensurerte levetidene.

Vi antar minst mulig om sensureringen, men vi antar at den er uavhengig. Hvorvidt en person forsvinner fra undersøkelsen må ikke ha noen sammenheng med det vi studerer. Hvis vi studerer virkningen av en vaksine mot alzheimer vil kanskje mange av forsøkspersonene bli borte og dermed sensurert fordi de har fått nettopp alzheimer og glemmer å møte opp på oppfølgingsmøtene. Vi antar at vi har de usensurerte, uavhengige levetidene  $T_1, \dots, T_n$ , og at  $\alpha_i(t)$  er hasardraten til individ  $i$ . Som nevnt tidlige kommer vi høyst sannsynlig ikke til å observere alle de faktiske levetidene. Noen overlever til undersøkelsen er over, mens andre vil bli sensurert av andre grunner. Det vi observerer er en sensurert levetid  $\tilde{T}_i$  sammen med en indikator  $D_i$ , der  $D_i = 1$  hvis  $\tilde{T}_i = T_i$  og  $D_i = 0$  hvis  $\tilde{T}_i < T_i$ . Det vil si at  $D_i = 1$  hvis vi observerer en faktisk levetid, mens  $D_i = 0$  hvis vi observerer en sensurert levetid.

Vi har nå uavhengig høyresensurering hvis en person som fortsatt ikke har opplevd hendelsen vi studerer eller er sensurert ved tid  $t$  har samme sannsynlighet for å oppleve hendelsen



i det neste korte tidsintervall  $[t, t+dt)$  som hvis vi ikke hadde hatt sensurering. Definisjonen på uavhengig sensurering er dermed:

$$P(t \leq \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i \geq t, \text{fortid}) = P(t \leq T_i < t + dt | T_i \geq t) \quad (2.3)$$

Med “fortid” menes her informasjon om hvorvidt  $\tilde{T}_i < t$  eller  $\tilde{T}_i \geq t$ , og verdien av  $D_i$  for alle  $i$  der  $\tilde{T}_i < t$ .

På samme måte som at forsøkspersonene kan gå ut av studien på grunn av død eller fordi de får sitt første barn eller av andre grunner, kommer de også ofte inn i studien ved forskjellige tidspunkter. Dette kalles trunkering. I en del medisinske studier vil for eksempel personene først komme inn når de er diagnostisert eller en bestemt hendelse har inntruffet.

Som nevnt er det ofte ikke forsøkspersonenes tid til død vi er ute etter. Døden er jo noe som inntreffer for alle før eller siden. En kvinnes første barn trenger derimot ikke å komme noen gang, og de fleste ekteskap ender ikke i skilsmisse. Vi kan derfor ikke ekskludere fra studien alle dem som ikke opplevde den hendelsen vi studerer. Da ville vi sitte igjen med data der alle forsøkspersonene føder minst ett barn, og der alle ekteskapene går i oppløsning. Vi må derfor ta med oss også denne informasjonen når vi skal analysere dataene.

En forsøksperson som fortsatt er med i undersøkelsen og derfor fortsatt kan oppleve den hendelsen vi interesserer oss for, sier vi er under risiko. Gruppen forsøkspersoner som er under risiko kaller vi risikomengden. Dersom en person dør, er han ikke under risiko for noe som helst lenger, og han faller også ut av risikomengden. I studien der vi var interesserte i tidspunktet for kvinners første barn, vil også kvinnene falle ut av risikomengden med en gang de er blitt mødre. De personene som sensureres av andre årsaker faller også ut av risikomengden i likhet med dem som ikke lenger kan oppleve den hendelsen vi er ute etter.

### 2.1.5 Gjentatte hendelser

Vi skiller mellom hendelser som bare kan skje en gang og hendelser som kan skje flere ganger. Eksempler på det sistnevnte kan være fødsler og beinbrudd, mens et eksempel på hendelser som bare kan skje en gang er død.

Når vi opererer med gjentatte hendelser blir noen ting forskjellige fra i overlevelsesanalysen. Et individ kan være under risiko selv om hendelsen allerede har inntruffet for dette individet. Typisk i disse situasjonene er at enkelte individer kan oppleve mange hendelser, mens andre ikke opplever hendelsen i det hele tatt. Vi vil se lite på situasjoner med gjentatte hendelser, men heller konsentrere oss om overlevelsesanalyse.

### 2.1.6 Statistiske metoder for kohortdata

Den kumulative hasarden viser seg å være lettere å estimere enn hasardraten. Gitt at vi har de sorterte observerte levetidene  $t_1 < t_2 < \dots$ , det vil si  $\tilde{T}_i$ -ene med  $D_i = 1$ , kan vi

estimerer den kumulative hasarden med Nelson-Aalen estimatoren. Den er gitt som

$$\hat{A}(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)},$$

der  $Y(t_j)$  er antall individer i risikomengden ved tid  $t_j$ , dvs antall individer under risiko rett før tid  $t_j$ . Det kan vises at variansen til Nelson-Aalen estimatoren kan estimeres ved

$$\hat{\sigma}^2(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)^2}.$$

Nelson-Aalen estimatoren er tilnærmet normalfordelt i et gitt tidspunkt  $t$ , når vi har store utvalg.

**Eksempel 2 (Nelson-Aalen) ??** Figur 2.1 viser kummulativ hasard til pasientene i eksempel 1 gruppert etter behandling, kjønn, ascites, alder og protrombin index. Stigningstallet til kurvene er hasardraten. Vi ser av plottene at behandlingen ikke ser ut til å være særlig effektiv. Det er heller ingen tydelig forskjell på kurvene til kvinner og menn. Derimot ser vi at det er tydelige forskjeller mellom de forskjellige aldersgruppene, avhengig av hvor mye ascites de hadde ved studiestart, og protrombin index. Disse forklaringsvariablene virker dermed å være viktige for å forklare eller forstå overlevelsen til pasientene.

Overlevelsesfunksjonen kan estimeres med Kaplan-Meier estimatoren. Denne er definert som

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{1}{Y(t_j)}\right)$$

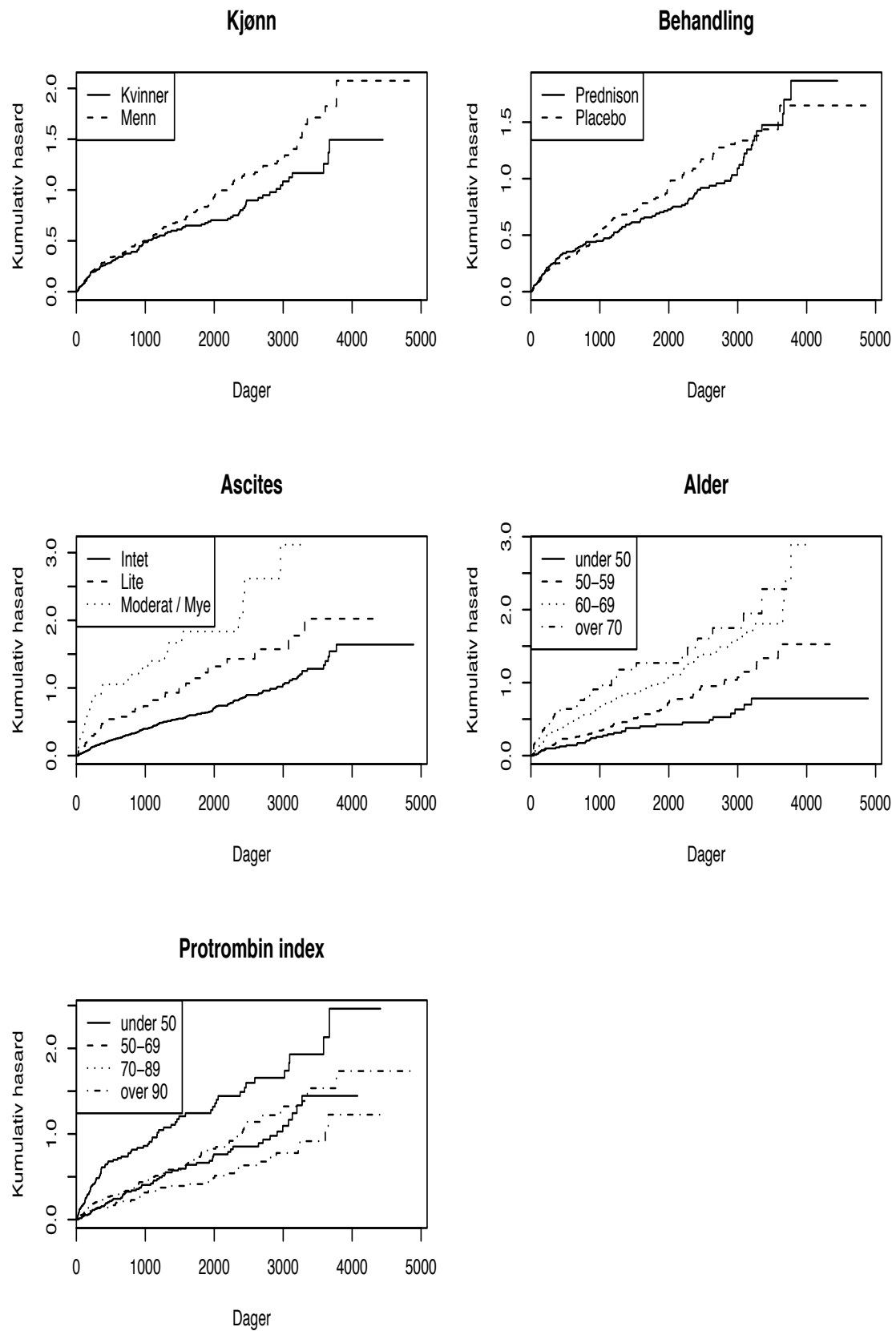
Variansen til Kaplan-Meier estimatoren kan estimeres med Greenwoods formel

$$\hat{\tau}^2(t) = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{1}{(Y(t_j)(Y(t_j) - 1))}.$$

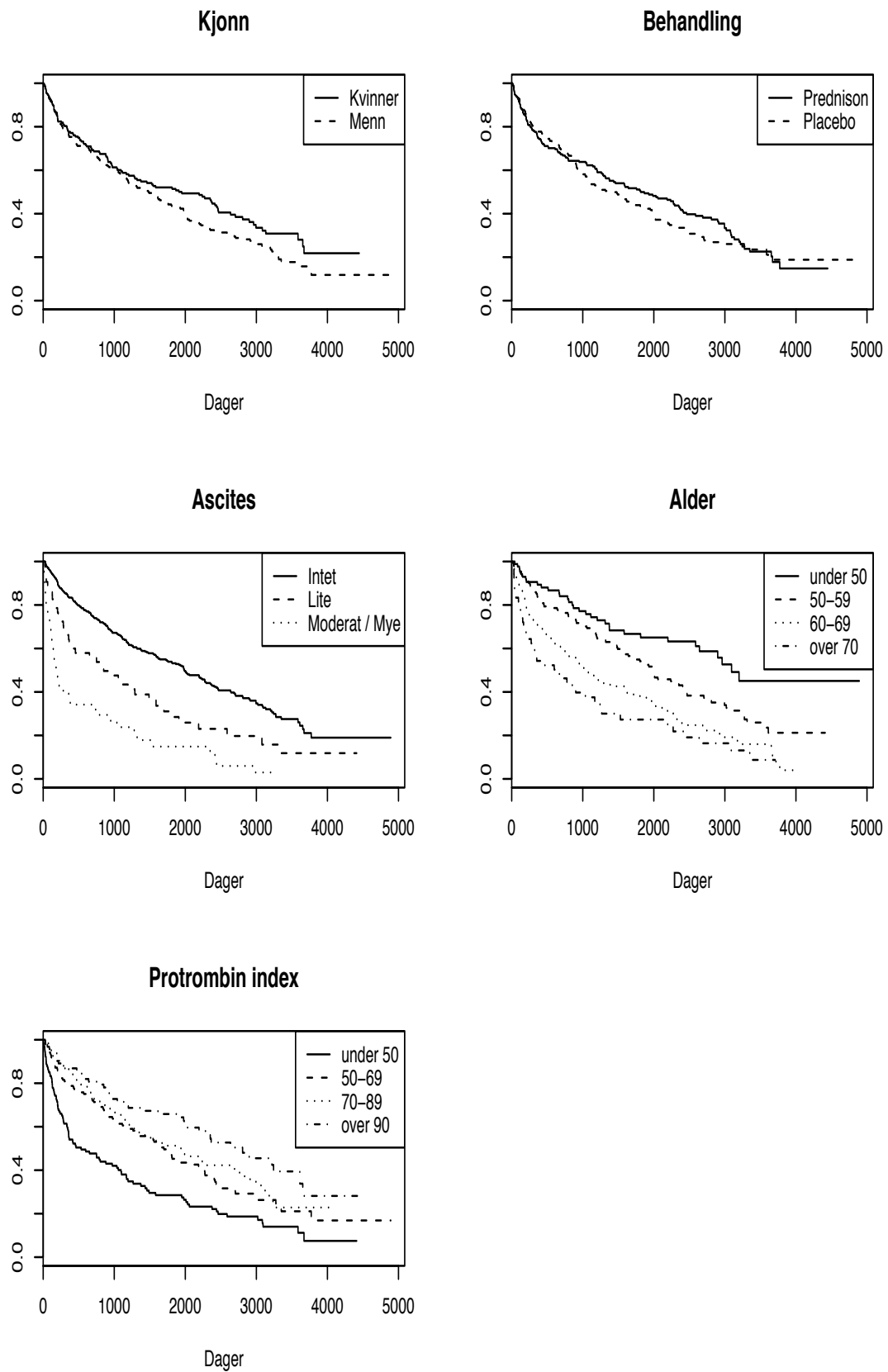
Det kan vises at Kaplan-Meier estimatoren er tilnærmet normalfordelt ved et gitt tidspunkt  $t$ , når vi har store utvalg.

**Eksempel 3 (Kaplan-Meier)** I figur 2.2 har vi laget Kaplan-Meier plott som viser overlevelsen til personene i skrumpleverundersøkelsen, gruppert etter de forskjellige kovariablene. På samme måte som av Nelson-Aalen plottet ser vi at det ikke er lett å se noen stor behandlingseffekt. Kvinner ser ut til å gjøre det noe bedre enn menn, men kjønn ser ikke ut til å være viktig for å forklare overlevelsen til forsøkspersonene. Alder, protrombin index og spesielt ascites har derimot store forskjeller mellom gruppene, og virker dermed å være viktige for overlevelsen. Personer med moderat eller mye ascites har median levetid rundt 100 dager, tilsvarende for dem uten ascites er rundt 2000 dager.

Til nå har vi sett på metoder som baserer seg på å dele populasjonen inn i to eller flere underpopulasjoner og så se på forskjellene mellom disse. Ofte har vi flere kovariater og



Figur 2.1: Nelson-Aalen plottene viser kumulativ hasard for kovarietene.



Figur 2.2: Kaplan-Meier. Kurvene viser observert kummulativ hasard for de ulike gruppe-  
ne.

noen er numeriske. I slike situasjoner trenger vi regresjonsmodeller. Data kan tilpasses ulike modeller for å estimere sammenhengen mellom hasardraten og kovariatene. Vi antar at vektoren av kovariater  $\mathbf{x}_i(t) = (x_{i1}(t), \dots, x_{ip}(t))^T$  for individ  $i$  til tid  $t$  er relatert til hasardraten  $\alpha(t|\mathbf{x}_i)$  til individet ved relasjonen:

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t)r(\boldsymbol{\beta}, \mathbf{x}_i(t)), \quad (2.4)$$

der  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^T$  er en vektor av regresjonsparametre og  $r(\boldsymbol{\beta}, \mathbf{x}_i(t))$  er relativ risiko funksjonen som er slik at  $r(0, x_i(t)) = 1$ . Baseline hasarden  $\alpha_0(t)$  er hasardraten til en person med alle kovariater lik 0. I Cox-regresjon er relativ risiko funksjonen

$$r(\boldsymbol{\beta}, \mathbf{x}_i(t)) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i(t)). \quad (2.5)$$

Hasardratioen til to personer blir da

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_2(t))}{\exp(\boldsymbol{\beta}^T \mathbf{x}_1(t))} = \exp(\boldsymbol{\beta}^T (\mathbf{x}_2(t) - \mathbf{x}_1(t))).$$

Hvis  $\mathbf{x}_1(t)$  og  $\mathbf{x}_2(t)$  er like bortsett fra at  $x_{2j}(t) = x_{1j}(t) + 1$  får vi at hasardratioen i Cox regresjon blir

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = e^{\beta_j}.$$

Tolkningen av  $\beta$ -ene er altså at dersom to individer har alle kovariater like, bortsett fra den  $j$ -te kovariatene der individ 2 har en økning på en enhet i forhold til individ 1, så vil individ 2 ha  $e^{\beta_j}$  ganger høyere hasard enn individ 1.

Cox-regresjon er en mye brukt modell som bygger på antakelsen om proposjonale hasarder dersom kovariatene ikke avhenger av tiden. Denne antakelsen innebærer at hasarden til individer forårsaket av ulike kovariater er proposjonale. Dersom røyking doubler risikoen for lungekreft ved et tidspunkt  $t$ , antar vi med andre ord at den vil doble risikoen ved alle  $t$ .

Koeffisientene i regresjonsmodellen estimeres ved å maksimere den partielle likelihood-funksjonen. Denne finner vi ved å multiplisere sammen de betingete sannsynlighetene for at det er individ  $i_j$  som opplever en hendelse i tid  $t_j$ , gitt at et individ har enn hendelse i tid  $t_j$  og fortiden. Den partielle likelihoodfunksjonen til  $\boldsymbol{\beta}$  blir i den generelle modellen

$$L(\boldsymbol{\beta}) = \prod_{t_j} \frac{r(\boldsymbol{\beta}, \mathbf{x}_{i_j}(t_j))}{\sum_{l \in R_j} r(\boldsymbol{\beta}^T \mathbf{x}_l(t_j))}, \quad (2.6)$$

der  $i_j$  individet som opplever en hendelse i tidspunkt  $t_j$  og  $\mathcal{R}_j$  er risikomengden ved tid  $t_j$ , det vil si  $\{l | Y_l(t_j) = 1\}$ . Vi finner da vektoren av de estimerte regresjonskoeffisientene  $\hat{\boldsymbol{\beta}}$  som den verdi av  $\boldsymbol{\beta}$  som maksimerer  $L(\boldsymbol{\beta})$ . Spesielt for Cox regresjon finner vi at den partielle likelihoodfunksjonen til  $\boldsymbol{\beta}$  blir

$$L(\boldsymbol{\beta}) = \prod_{t_j} \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_{i_j}(t_j))}{\sum_{l \in R_j} \exp(\boldsymbol{\beta}^T \mathbf{x}_l(t_j))}.$$

$\hat{\beta}$  er tilnæmet  $N(\beta, \Sigma)$ -fordelt. Kovariansmatrisen  $\Sigma$  estimeres som  $\mathcal{I}(\hat{\beta})^{-1}$  der

$$\mathcal{I}(\beta) = \left\{ -\frac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_k} \right\}$$

er den observerte informasjonsmatrisen. Kumulativ baseline hasard  $A_0(t) = \int_0^t \alpha_0(u) du$  kan estimeres med

$$\hat{A}_0(t) = \sum_{t_j \leq t} \frac{1}{\sum_{l \in \mathcal{R}_j} r(\hat{\beta}, \mathbf{x}_l(t_j))} \quad (2.7)$$

som kalles Breslowestimatoren.

For å teste nullhypotesen  $\beta = \beta_0$  kan følgende tester brukes:

- Wald-testen

$$\chi_W^2 = (\hat{\beta} - \beta_0)^T \mathcal{I}(\hat{\beta}) (\hat{\beta} - \beta_0)$$

- Likelihood ratio testen

$$\chi_{LR}^2 = 2(\log L(\hat{\beta}) - \log L(\beta_0))$$

- Score-testen

$$\chi_{SC}^2 = \mathbf{U}(\beta_0)^T \mathcal{I}(\beta_0)^{-1} \mathbf{U}(\beta_0)$$

der  $\mathbf{U}(\beta) = \frac{\partial}{\partial \beta} \log L(\beta)$  er vektoren av score-funksjoner.

Disse testobservatorene er asymptotisk like og  $\chi_p^2$ -fordelt under  $H_0$ .

**Eksempel 4 (Cox-regresjon)** I tabell 2.2 ser vi resultatet etter å ha tilpasset Cox-regresjonsmodellen til skrumpleverdataene. Ascites og protrombin index er faktorisert. Grensene for protrombin index er valgt som i eksempel 1.

Tabell 2.2: Resultatene av Cox regresjon på skrumpleverdataene.

Kovariat	$\hat{\beta}$	$e^{\hat{\beta}}$	$se(\hat{\beta})$	$z$	$p$
Behandling	0.104	1.109	0.119	0.87	$3.8 \cdot 10^{-1}$
Kjønn	0.557	1.746	0.128	4.34	$1.4 \cdot 10^{-5}$
Ascites1	0.431	1.539	0.179	2.41	$1.6 \cdot 10^{-2}$
Ascites2	1.018	2.768	0.180	5.64	$1.7 \cdot 10^{-8}$
Alder	0.050	1.051	0.007	7.33	$2.4 \cdot 10^{-13}$
protrombin index (50,69]	-0.524	0.592	0.152	-3.46	$5.4 \cdot 10^{-4}$
protrombin index (69,89]	-0.745	0.475	0.175	-4.25	$2.1 \cdot 10^{-5}$
protrombin index (89, -]	-0.811	0.444	0.185	-4.39	$1.1 \cdot 10^{-5}$

Vi ser av tabellen at for eksempel dersom to personer er helt like på alle måter, bortsett fra at den ene er ett år eldre enn den andre, så har den eldste 1.05 ganger så høy hasard ved en gitt tid som den yngste.

Vi husker fra Nelson-Aalen og Kaplan-Meier plottene at behandlingen ikke så ut til å være særlig effektiv. Derfor er det ikke så overaskende at behandlingen ikke blir en signifikant kovariat i Cox-regresjonsmodellen. Dette betyr at behandling ikke ser ut til å være en viktig forklaringsvariabel for å forstå eller forklare overlevelsen til pasientene. Alle de andre kovariatene er derimot meget signifikante, med p-verdi  $< 0.01$ . En mer grundig analyse er nødvendig for å forklare disse sammenhengene.

## 2.2 Telleprosess formulering av kohort data

Telleprosesser er et nyttig verktøy for å formulere og analysere modeller for levetider. Vi ser på en kohort med  $n$  individer. La  $N_i(t)$  være antall hendelser som har inntruffet til og med tid  $t$  for individ  $i$ .  $N_i(t)$  er da en telleprosess. I overlevelsesanalysen kan telleprosessen til et individ bare ta verdiene 0 og 1. I situasjonen med gjentatte hendelser vil telleprosessen også kunne ta høyere verdier.

Inkrementet til telleprosessen,  $dN_i(t)$ , altså forandringen i  $N_i(t)$  i det korte intervallet  $[t, t + dt)$ , er

$$dN_i(t) = \begin{cases} 1 & \text{hvis en hendelse inntreffer for individ } i \text{ i } [t, t + dt) \\ 0 & \text{ellers} \end{cases}$$

Vi ser nå på alle de  $n$  individene, og lar som over  $t_j$  være tidspunktet da den  $j$ -te hendelsen inntraff, mens  $i_j$  er individet som hadde en hendelse ved tid  $t_j$ . Da er telleprosessen til individ  $i$  gitt som

$$N_i(t) = \sum_{t_j \leq t} I(i_j = i) = I(\tilde{T}_i \leq t, D_i = 1),$$

$i = 1, 2, \dots, n$ . Vi husker at  $\tilde{T}_i$  er den sensurerte levetiden til individ  $i$ , mens  $D_i$  er en indikator for hvorvidt vi faktisk observerte levetiden til individ  $i$ , eller om den ble sensurert. Vi antar for hvert individ at to hendelser aldri inntreffer samtidig. Dette antar vi også at gjelder for kohorten. Telleprosessen til kohorten teller opp antall hendelser som har inntruffet i kohorten til og med tid  $t$ . Denne er summen av de individuelle telleprosessene

$$N(t) = \sum_{i=1}^n N_i(t) = \sum_{i=1}^n I\{\tilde{T}_i \leq t, D_i = 1\}.$$

Intensitetsprosessen  $\lambda_i(t)$  til en telleprosess  $N_i(t)$ , er den betingede sannsynligheten gitt fortiden for at en hendelse inntreffer for individ  $i$  i det korte intervallet  $[t, t + dt)$ , delt på lengden til intervallet. Med fortiden mener vi alt som er observert før tidspunkt  $t$ , og skriver  $\mathcal{H}_{t-}$ . Intensitetsprosessen til individ  $i$  blir dermed:

$$\lambda_i(t)dt = P(dN_i(t) = 1 | \mathcal{H}_{t-}) \quad (2.8)$$

For overlevelsesdata blir dette sannsynligheten for at vi observerer at individ  $i$  dør i neste korte intervall, gitt at han levde og ikke var sensurert ved intervallet begynnelse.

Antagelsen om uavhengig sensurering (2.3) kan uttrykkes ved hjelp av intensitetsprosessen til en telleprosess. Vi får at

$$\lambda_i(t)dt = P(dN_i(t) = 1|\mathcal{H}_{t-}) = P(t \leq \tilde{T}_i < t + dt, D_i = 1|\mathcal{H}_{t-}).$$

Siden

$$\lambda_i(t)dt = P(dN_i(t) = 1|\mathcal{H}_{t-}) = 0$$

når  $\tilde{T}_i < t$  og

$$\lambda_i(t)dt = P(dN_i(t) = 1|\mathcal{H}_{t-}) = P(t \leq T_i < t + dt|T_i \geq t) = \alpha_i(t)dt$$

når  $\tilde{T}_i > t$ , får vi at uavhengig sensurering medfører at intensitetsprosessen til en telleprosess, gitt at hendelsen vi observerer ikke har inntruffet enda, blir lik hasardraten,  $\alpha_i(t)$  til individet. Når hendelsen har inntruffet er intensitetsprosessen lik 0. Med uavhengig sensurering tar intensitetsprosessen til individ  $i$  dermed formen

$$\lambda_i(t) = \alpha_i(t)Y_i(t),$$

der  $Y_i(t) = I\{\tilde{T}_i \geq t\}$  er en indikator for om individ  $i$  er under risiko.  $Y_i(t) = 0$  hvis individ  $i$  ikke er under risiko ved tid  $t$ , mens den er 1 hvis individet er under risiko ved tid  $t$ .

Den kumulative intensitetsprosessen til individ  $i$  er gitt ved

$$\Lambda_i(t) = \int_0^t \lambda_i(u)du = \int_0^t \alpha_i(u)Y_i(u)du,$$

Vi ser nå på forskjellen på telleprosessen og den kumulative intensitetsprosessen, for individ  $i$ :

$$\begin{aligned} M_i(t) &= N_i(t) - \Lambda_i(t) \\ dM_i(t) &= dN_i(t) - \lambda_i(t)dt \end{aligned} \tag{2.9}$$

Siden  $dN_i(t)$  er en binær variabel kan vi skrive (2.8) som

$$\lambda_i(t)dt = E(dN_i(t)|\mathcal{H}_{t-}).$$

Siden  $\lambda_i(t)dt$  er en funksjon av  $\mathcal{H}_{t-}$  får vi at

$$E(dN_i(t) - \lambda_i(t)dt|\mathcal{H}_{t-}) = 0.$$

Av (2.9) får vi dermed at

$$E(dM_i(t)|\mathcal{H}_{t-}) = 0$$

som betyr at  $M_i(t)$  er en martingal.

Det følger også av (2.9) at

$$dN_i(t) = \lambda_i(t)dt + dM_i(t).$$

Dette er et uttrykk på formen “observasjon” = “modell” + “støy”, der telleprosessen er observasjon, intensitetsprosessen er modell og martingale residualprosessen er støy. Tolkningen til  $dM_i(t)$  blir altså at den er støyledet, mens  $M_i(t)$  er kumulativ støy.



### 2.2.1 Når modellen stemmer

Vi husker at hasardraten til et individ  $i$  var definert som den betingete sannsynligheten for at individet vil oppleve en hendelse i det neste korte intervallet  $[t, t + dt)$ , gitt at hendelsen ikke har inntruffet før tid  $t$ , altså

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T_i < t + dt | T_i \geq t) = \alpha_i(t).$$

Hvis vi har uavhengig sensurering har vi at intensitetsprosessen til individ  $i$  kan skrives på formen

$$\lambda_i(t) = \alpha_i(t) Y_i(t).$$

Slik er det i virkeligheten og vi ønsker å modellere dette. Dette gjør vi ved å tilpasse en modell til data. Det finnes en rekke ulike modeller, men vi vil foreløpig se på relativ risiko regresjonsmodeller generelt. For å tilpasse modeller må vi gjøre visse antagelser som avhenger av hvilken regresjonsmodell vi bruker. Hvis regresjonsmodellen stemmer med virkeligheten har vi at (2.4) holder

$$\alpha_i(t) = \alpha_0(t) r(\beta, x_i(t)),$$

det vil si at modellantagelsen holder. Intensitetsprosessen blir dermed

$$\lambda_i(t) = Y_i(t) r(\beta, x_i(t)) \alpha_0(t),$$

mens den kumulative intensitetsprosessen blir

$$\Lambda_i(t) = \int_0^t Y_u(t) r(\beta, x_i(u)) \alpha_0(u) du$$

der  $\alpha_0(u) du = dA_0(u)$ . Vi kan estimere  $\Lambda_i(t)$  ved å sette inn maximum likelihood estimatet  $\hat{\beta}$  for  $\beta$ , og bruke inkrementet til Breslowestimatoren fra (2.7) for  $\alpha_0(u) du$ .

Ved å sette inn estimatet  $\hat{\Lambda}_i(t)$  for  $\Lambda_i(t)$  kan vi finne et uttrykk for den estimerte martingal residualprosessen:

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) r(\hat{\beta}, \mathbf{x}_i(u)) d\hat{A}_0(u)$$

Hvis  $\tau$  er studiens siste tidspunkt blir martingal residualene

$$\hat{M}_i = \hat{M}_i(\tau) = N_i(\tau) - \hat{\Lambda}_i(\tau).$$

Disse estimerte martingal residual prosessene vil spille en viktig rolle i evaluering av modelltilpasning senere i oppgaven. I neste kapittel vil vi introdusere plot av estimerte martingal residual prosesser og en formell goodness-of-fit test som baserer seg på estimerte martingal residual prosesser som verktøy for å studere modelltilpasning.



## Kapittel 3

# Kasus-kontroll studier

Hittil har vi studert metoder egnet for å analysere kohortdata der informasjon om hvert enkelt individ i studien er tilgjengelig. Ofte er sykdommen vi studerer sjelden, og dermed må svært mange individer følges opp for at vi skal registrere et tilstrekkelig antall hendelser. Da vil det være dyrt og tidkrevende å fremskaffe informasjon om alle individene i kohorten. I slike situasjoner er metoder som baserer seg på et mindre antall kontroller ofte å foretrekke. Særlig gjelder dette altså studier med stor kohort og hvor informasjonen om individene ikke er tilgjengelig på forhånd. Vi vil i dette kapitlet introdusere to ulike samplingdesign for kasus-kontroll studier, der den første baserer seg på å trekke et tilfeldig utvalg kontroller, mens den andre metoden nyttegjør eventuell tilgjengelig informasjon om individene i kontrollgruppen til å trekke et mer informativt utvalg kontroller.

Videre vil vi vise hvordan vi kan benytte telleprosesser til å analysere disse forsøksdesignene. Med dette verktøyet på plass kan vi introdusere grupperte martingal residualprosesser som et nyttig verktøy for å undersøke modelltilpasningen. Plot av martingal residualprosessene samt mer formelle tester vil brukes for å studere modelltilpasning

Stoffet i dette kapitlet er hentet fra Borgan og Langholz (2007), og Borgan, Goldstein og Langholz (1995).

### 3.1 Innledning

Hittil har vi sett på kohortstudier med proposjonale hasardmodeller. Disse baserer seg på at kovariatinformasjonen til individer som dør eller opplever en hendelse av interesse, sammenlignes med forklaringsvariablene til individene som ikke har dødd eller opplevd hendelsen av interesse (enda). Vi ser av den partielle likelihooden til regresjonskoeffisientene  $\beta$  for kohortdata (2.6) at vi ved hver hendelsestid summerer over alle individer under risiko. Vi trenger derfor kovariatinformasjonen til alle individene. Dersom begivenheten som studeres er sjelden, betyr dette vanligvis at kohortstudier krever at et høyt antall individer følges opp i en lengre periode, for at tilstrekkelig mange dødsfall skal inntreffe slik at studien får stor nok statistisk kraft til å gi pålitelige resultater. Slike studier blir

dermed dyre og vanskelig å gjennomføre. Det er også unødvendig å bruke mye ressurser på slike undersøkelser i situasjoner der man kan oppnå omtrent like gode resultater med et annet design. Hvis sykdommen som studeres er sjelden, vil individene som overlever bidra svært lite til studiens statistiske kraft i forhold til individene som dør. Å bruke ressurser på å samle inn kovariatinformasjon fra alle individene i kohorten vil derfor være unødvendig.

Et alternativ til kohortstudier er studiedesign som baserer seg på å samle inn kovariatinformasjon fra alle individene som dør i studieperioden, men bare et mindre antall kontrollpersoner. Dette reduserer ressursbruken knyttet til å samle inn data, uten nødvendigvis å gå særlig på bekostning av resultatene. En annen motivasjon for å bruke slike design er at med en gitt ressursmengde kan mer presise resultater oppnås ved å sample fra en større kohort, enn å undersøke alle individene i en mindre kohort. Vi vil i dette kapitlet se nærmere på slike design.

## 3.2 Sampling

I kasus-kontroll studier velges ved hvert hendelsestidspunkt  $t_j$  en samlet risikomengde  $\mathcal{R}_j$  som består av individet  $i_j$  som døde ved tid  $t_j$  samt et gitt antall kontrollpersoner. Kontrollpersonene velges blant de  $Y(t_j) - 1$  individene som er under risiko, men ikke opplever hendelsen ved tid  $t_j$ . Forskjellige design baserer seg på ulike samplingstrategier av kontrollene. Vi vil i det følgende se nærmere på enkel nøstet kasus-kontroll sampling og counter-matching.

Før vi ser nærmere på samplingen av kontroller må vi introdusere noen størrelser. Kohort og samplinghistorien  $\mathcal{F}_{t-}$  inneholder informasjon om hendelsene,  $\mathcal{H}_{t-}$ , samt sampling av kontrollene inntil, men ikke i, tid  $t$ . La  $\mathcal{P}$  være potensmengden av  $1, 2, \dots, n$ , det vil si mengden av alle delmengder av  $\{1, 2, \dots, n\}$ . La  $\mathcal{P}_i = \{\mathbf{r} : \mathbf{r} \in \mathcal{P}, i \in \mathbf{r}\}$ , altså alle mengder i  $\mathcal{P}$  som også inneholder individ  $i$ . I vår kontekst inneholder  $\mathcal{P}$  alle mulige samplede risikomengder, mens  $\mathcal{P}_i$  inneholder alle mulige samplede risikomengder der individ  $i$  inngår.

Gitt  $\mathcal{F}_{t-}$ , hvis individ  $i$  dør ved tid  $t$ , velges mengden  $\mathbf{r} \in \mathcal{P}_i$  som samlet risikomengde med sannsynlighet  $\pi_t(\mathbf{r}|i)$ . Vi definerer at  $\pi_t(\mathbf{r}|i) = 0$  når  $Y_i(t) = 0$ . For et individ  $i$  som er under risiko,  $Y_i(t) = 1$ , blir sannsynligheten til de mulige risikomengdene, gitt at individ  $i$  dør ved tid  $t$ ,  $\pi_t(\mathbf{r}|i)$ , en sannsynlighetsfordeling over mengdene  $\mathbf{r} \in \mathcal{P}_i$ .

Vi vil introdusere noen faktoriseringer som vil bli nyttige senere. Først setter vi

$$\pi_t(\mathbf{r}) = Y(t)^{-1} \sum_{l \in \mathbf{r}} \pi_t(\mathbf{r}|l)$$

Vi finner altså  $\pi_t(\mathbf{r})$  ved å summere over de individuelle, betingede sannsynlighetene for å velge risikomengde  $\mathbf{r}$ , og dele på antall under risiko. Videre ser vi at  $\pi_t(\mathbf{r})$  er en sannsynlighetsfordeling over alle mengdene  $\mathbf{r} \in \mathcal{P}$ :

$$\sum_{\mathbf{r} \in \mathcal{P}} \pi_t(\mathbf{r}) = Y(t)^{-1} \sum_{l=1}^n \sum_{\mathbf{r} \in \mathcal{P}_i} \pi_t(\mathbf{r}|l) = Y(t)^{-1} \sum_{l=1}^n Y_l(t) = 1$$

siden sannsynlighetene for å velge en av de mulige risikomengdene i  $\mathcal{P}$  summerer til én. Vi introduserer videre vektene

$$w_i(t, \mathbf{r}) = \frac{\pi_t(\mathbf{r}|i)}{\pi_t(\mathbf{r})}.$$

Vektene avhenger av samplingstrategien, og vi vil komme tilbake til disse når vi tar for oss de ulike strategiene for kontrollsampling. Videre har vi faktoriseringen

$$\pi_t(\mathbf{r}|i) = w_i(t, \mathbf{r})\pi_t(\mathbf{r}) \quad (3.1)$$

Samplingsannsynligheten  $\pi_t(\mathbf{r}|i)$  kan avhenge av hendelser som har inntruffet i fortiden,  $\mathcal{F}_{t-}$ , men ikke hendelser som vil inntreffe i fremtid. Individer som dør ved et senere tidspunkt blir for eksempel ikke ekskludert fra den samlede risikomengden ved tidligere tidspunkter. Samplingen skjer uavhengig ved hvert tidspunkt slik at et individ kan inngå i den samlede risikomengden ved flere tidspunkter.

På samme måte som vi antar uavhengig sensurering og trunkering, antar vi uavhengig sampling av risikomengder. Informasjon om hvilke individer som har blitt med i undersøkelsen, blitt sensurert eller blitt valgt som kontroller før tid  $t$  inneholder ingen informasjon om dødsrisikoen ved tid  $t$ . I kapittel 3.5 vil vi introdusere telleprosesser og intensitetsprosesser, og vil da si at vi har uavhengig sampling hvis  $(\mathcal{F}_{t-})$ -intensitetsprosessene til telleprosessene er den samme som  $(\mathcal{H}_{t-})$ -intensitetsprosessene.

### 3.3 Nøstet kasus-kontroll design

I en kasus-kontroll studie sammenlignes ved hvert hendelsestidspunkt  $t$  kovariatinformasjonen til kasus  $i_j$  med kovariatinformasjonen til et utvalg kontroller. Dette utvalget av kontroller utgjør sammen med kasus ved tidspunktet den samlede risikomengden  $\tilde{\mathcal{R}}_j$ . Kontrollene kan samples på ulike måter som gir opphav til forskjellige forsøksdesign. Vi vil her se nærmere på to av disse. Enkel nøstet kasus-kontroll er den enkleste designen med tilfeldig sampling av kontroller blant individene under risiko. Counter-matching nyttegjør seg eventuell informasjon som er tilgjengelig om individene i kohorten til å gruppere individene, og dermed sample en mer informativ risikomengde.

#### 3.3.1 Enkel nøstet kasus-kontroll design

I en enkel nøstet kasus-kontroll design velges ved hver hendelsestid  $t$ ,  $m - 1$  kontroller til den samlede risikomengden, sammen med individet som dør ved tid  $t$ . Antallet  $m$  er her gitt. De  $m - 1$  kontrollene velges ved tilfeldig utvalg uten tilbakelegging blant de  $Y(t) - 1$  individene under risiko ved tid  $t$ . Sannsynligheten for å velge en samlet risikomengde  $\mathbf{r}$  ved hendelsestid  $t$ , gitt at individ  $i$  dør, blir nå

$$\pi_t(\mathbf{r}|i) = \left( \frac{Y(t) - 1}{m - 1} \right)^{-1} I\{|\mathbf{r}| = m, \mathbf{r} \in \mathcal{R}(t)\}$$

for alle mulige risikomengder som inneholder individ  $i$ , det vil si alle  $\mathbf{r} \in \mathcal{P}_i$ . Vi finner videre sannsynligheten for å velge risikomengde  $\mathbf{r}$  som

$$\pi_t(\mathbf{r}) = \left( \frac{Y(t)}{m} \right)^{-1} I\{|\mathbf{r}| = m, \mathbf{r} \in \mathcal{R}(t)\}; \quad \mathbf{r} \in \mathcal{P};$$

og vektene til individ  $i$  blir

$$w_i(t, \mathbf{r}) = \frac{Y(t)}{m} I\{i \in \mathbf{r}\}. \quad (3.2)$$

Merk at vektene ved enkel nøstet kasus-kontroll blir like for alle de  $m - 1$  kontrollene som samples ved hvert hendelsestidspunkt, samt kasuset.

### 3.3.2 Counter-matching

For enkel nøstet kasus-kontroll trenger vi bare å vite hvem som er under risiko når vi trekker kontrollene. Eventuell informasjon om individene nyttes ikke i trekkingen av kontrollene. Ofte er noe informasjon tilgjengelig eller lett tilgjengelig for alle individene. Dette kan være alder, jobb, kjønn etc. Langholz og Borgan (1995) laget en stratifisert versjon av enkel nøstet kasus kontroll som benytter seg av den tilgjengelige informasjonen i trekkingen av kontrollmengde. Hensikten er å få en mer informativ samplet risikomengde.

I counter-matching deles individene inn i  $S$  strata på bakgrunn av den tilgjengelige informasjonen. Vi lar  $\mathcal{R}_s(t)$  betegne delmengden av risikomengden  $\mathcal{R}(t)$  som tilhører stratum  $s$ , mens  $Y_s(t) = |\mathcal{R}_s(t)|$  er antallet under risiko i stratum  $s$  rett før tid  $t$ . Ved dødsfall i tid  $t$  samples kontroller slik at den samplede risikomengden  $\tilde{\mathcal{R}}(t)$  består av  $m_s$  individer fra hvert stratum  $s = 1, \dots, S$ .

Dersom individ  $i$  fra stratum  $s(i)$  dør ved tid  $t$  trekkes tilfeldig  $m_s$  kontroller uten tilbakelegging fra hvert stratum  $s \neq s(i)$ . Fra stratum  $s(i)$  trekkes  $m_{s(i)} - 1$  kontroller på samme måte. Individ  $i$  inngår også i den samplede risikomengden  $\tilde{\mathcal{R}}(t)$  som dermed inneholder  $m_s$  individer fra hvert stratum. Stratifiseringen kan være tidsavhengig, men informasjonen som brukes for stratifisering ved tid  $t$  må være tilgjengelig rett før tid  $t$ . Vi kan dermed la individene bytte stratum dersom de for eksempel slutter å røyke eller bytter jobb.

For en gitt mengde  $\mathbf{r} \in \mathcal{P}_i$  som er en delmengde av  $\mathcal{R}(t)$  og tilfredstiller  $|\mathbf{r} \cap \mathcal{R}_s(t)| = m_s$ ,  $s = 1, \dots, S$ , har vi

$$\begin{aligned} \pi_t(\mathbf{r}|i) &= \left\{ \left( \frac{Y_{s(i)}(t) - 1}{m_{s(i)} - 1} \right) \prod_{s \neq s(i)} \left( \frac{m_s(t)}{m_s} \right) \right\}^{-1} \\ &= \left\{ \prod_{s=1}^S \left( \frac{Y_s(t)}{m_s} \right) \right\}^{-1} \frac{Y_{s(i)}(t)}{m_{s(i)}} \end{aligned}$$

Sannsynligheten for å velge en risikomengde  $\mathbf{r}$  blir ved counter-matching

$$\pi_t(\mathbf{r}) = \left\{ \prod_{s=1}^S \left( \frac{Y_s(t)}{m_s} \right) \right\}^{-1} I(|\mathbf{r} \cap \mathcal{R}_s(t)| = m_s, s = 1, \dots, S); \quad \mathbf{r} \in \mathcal{P};$$

og vektene for individ  $i$  blir

$$w_i(t, \mathbf{r}) = \frac{Y_{s(i)}(t)}{m_{s(i)}} I\{i \in \mathbf{r}\}. \quad (3.3)$$

Forskjellen på counter-matching og enkel nøstet kasus kontroll er at vi med counter-matching kan sikre stor variasjon i kovariatverdiene i den samplete risikomengden. Med enkel nøstet kasus-kontroll er det derimot tilfeldig hvilke individer som vil trekkes til den samplete risikomengden. Med større variasjon i kovariatverdiene får vi mer presise resultater.

### 3.4 Statistiske metoder for kasus-kontroll

Tidligere har vi vist at vi kan finne de estimerte regresjonskoeffisientene  $\hat{\beta}$  som verdien av  $\beta$  som maksimerer den partielle likelihood (2.6) for kohortdata. I tilfellet med kasus-kontroll finner vi regresjonskoeffisientene ved å maksimere den partielle likelihood for kasus-kontroll data generelt, gitt som

$$L(\beta) = \prod_{t_j} \frac{r(\beta, \mathbf{x}_{i_j}(t_j)) w_{i_j}(t_j, \tilde{\mathcal{R}}_j)}{\sum_{l \in \tilde{\mathcal{R}}_j} r(\beta, \mathbf{x}_l(t_j)) w_l(t_j, \tilde{\mathcal{R}}_j)}, \quad (3.4)$$

der  $i_j$  er kasuset til tid  $t_j$ , mens  $\tilde{\mathcal{R}}_j$  er den tilhørende samplete risikomengden. Vektene til individ  $l$ ,  $w_l(t_j, \tilde{\mathcal{R}}_j)$  bestemmes av samplingdesignet.

I tilfellet med enkel nøstet kasus-kontroll blir vektene  $w_{i_j}(t_j, \tilde{\mathcal{R}}_j)$ , gitt ved (3.2), like for individene under risiko ved tid  $t$ , og forkortes dermed. Den partielle likelihooden for enkel nøstet kasus kontroll blir dermed

$$L(\beta) = \prod_{t_j} \frac{r(\beta, \mathbf{x}_{i_j}(t_j))}{\sum_{l \in \tilde{\mathcal{R}}_j} r(\beta, \mathbf{x}_l(t_j))}.$$

For counter-matching blir likelihooden som (3.4), der vektene til hvert individ er gitt ved (3.3).

Den partielle likelihood for kohortdata (2.6) er et spesialtilfelle av (3.4). Hele risikomengden samples da med sannsynlighet én, og dermed blir  $\tilde{\mathcal{R}}_j = \mathcal{R}_j$  og vektene  $w_i(t, \mathbf{r}) = \frac{\pi_t(\mathbf{r}|i)}{\pi_t(\mathbf{r})} = \frac{1}{1} = 1$ . Vektene forkortes dermed.

Maximum partial likelihood estimatoren  $\hat{\beta}$  har de samme egenskapene i store utvalg som andre maximum likelihood estimatorer. Den er tilnærmet multinormalfordelt med forventning den sanne verdien av  $\beta$  som vi kaller  $\beta_0$ , og estimert kovariansmatrise  $I(\hat{\beta})^{-1}$ , der  $I(\beta)$  er den forventede informasjonsmatrisen, gitt som

$$I(\beta) = \sum_{t_j} \left\{ \frac{S_{\tilde{\mathcal{R}}_j}^{(2)}(\beta, t_j)}{S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j)} - \left( \frac{S_{\tilde{\mathcal{R}}_j}^{(1)}(\beta, t_j)}{S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j)} \right)^{\otimes 2} \right\}. \quad (3.5)$$

Her er

$$S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j) = \sum_{l \in \tilde{\mathcal{R}}_j} r(\beta, \mathbf{x}_l(t_j)) w_l(t_j, \tilde{\mathcal{R}}_j), \quad (3.6)$$

$$S_{\tilde{\mathcal{R}}_j}^{(1)}(\beta, t_j) = \sum_{l \in \tilde{\mathcal{R}}_j} \dot{r}(\beta, \mathbf{x}_l(t_j)) w_l(t_j, \tilde{\mathcal{R}}_j), \quad (3.7)$$

$$S_{\tilde{\mathcal{R}}_j}^{(2)}(\beta, t_j) = \sum_{l \in \tilde{\mathcal{R}}_j} \frac{\dot{r}(\beta, \mathbf{x}_l(t_j))^{\otimes 2}}{r(\beta, \mathbf{x}_l(t_j))} w_l(t_j, \tilde{\mathcal{R}}_j), \quad (3.8)$$

der  $\dot{r}(\beta, \mathbf{x}_l(t_j)) = \frac{\partial r(\beta, \mathbf{x}_l(t_j))}{\partial \beta}$ , og  $\mathbf{v}^{\otimes 2}$  til en kolonnevektor  $\mathbf{v}$  er matrisen  $\mathbf{v}\mathbf{v}^T$ .

Regresjonskoeffisientene  $\hat{\beta}$  velges som nevnt som den verdi av  $\beta$  som maksimerer den partielle likelihood til dataene. Det er det samme som å finne verdien som maksimerer den logaritmiske likelihood  $\log L(\beta)$ . Den logaritmiske likelihood til den generelle kasus-kontroll likelihooden (3.4) blir

$$l(\beta) = \sum_{t_j} \left\{ \log(r(\beta, \mathbf{x}_{i_j}(t_j))) + \log(w_{i_j}(t_j, \tilde{\mathcal{R}}_j)) - \log\left(\sum_{l \in \tilde{\mathcal{R}}_j} r(\beta, \mathbf{x}_l(t_j)) w_l(t_j, \tilde{\mathcal{R}}_j)\right) \right\}.$$

Dette kan skrives som

$$l(\beta) = \sum_{t_j} \left\{ \log(r(\beta, \mathbf{x}_{i_j}(t_j))) + \log(w_{i_j}(t_j, \tilde{\mathcal{R}}_j)) - \log(S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j)) \right\},$$

der  $S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j)$  er gitt ved (3.6). Nå kan vi finne score-vektoren, kolonnevektoren med de partielt deriverte mhp.  $\beta_1, \beta_2, \dots, \beta_p$  hhv, som

$$s(\beta) = \frac{\partial}{\partial \beta} l(\beta) = \sum_{t_j} \left\{ \frac{\dot{r}(\beta, \mathbf{x}_{i_j}(t_j))}{r(\beta, \mathbf{x}_{i_j}(t_j))} - \frac{S_{\tilde{\mathcal{R}}_j}^{(1)}(\beta, t_j)}{S_{\tilde{\mathcal{R}}_j}^{(0)}(\beta, t_j)} \right\} \quad (3.9)$$

Disse størrelsene er nødvendige i flere algoritmer for numerisk optimering av likelihood som vi vil bruke for å estimere regresjonskoeffisientene. Estimering av regresjonskoeffisientene må gjøres med en numerisk optimeringsmetode. Flere algoritmer er tilgjengelige. I denne oppgaven er Fishers scoring algoritme brukt. Denne er gitt som

$$\beta^{s+1} = \beta^s + I^{-1}(\beta^s) s(\beta^s), \quad (3.10)$$

der score-vektoren  $s(\beta^s)$  er gitt ved (3.9) mens  $I(\beta^s)$  er den forventede informasjonsmatrisen (3.5).

**Eksempel 5 (Numerisk estimering av regresjonskoeffisienter.)** I en studie av radoneksponering og røykings effekt på lungekreft, ble det mellom 1950 og 1960 rekruttert 3347 hvite urangruvearbeidere på Colorado Plataået i USA. Inntil 31. desember 1982 ble det fulgt opp om forsøkspersonene døde av lungekreft. Da hadde 258 dødsfall grunnet lungekreft blitt registrert. Informasjonen tilgjengelig for hver gruvarbeider inkluderer



eksponering for radon i working level months (WLM), og røykeeksponering. WLM er et mål på radoneksponering. Dette er konstruert slik at en person som har jobbet under en gitt stråling i en måned får samme score som en som har vært utsatt for dobbelt så høy stråling i en halv måned. Røykeeksponeringen er registrert i antall pakker. Radon og røykedataene analyseres begge som kumulative eksponering lagget to år, det vil si kumulativ eksponering inntil to år før tid  $t$ . Ved studier av lungekreft tenker man seg at eksponeringen individene har vært utsatt for den siste tiden før død har liten betydning for dødsrisikoen. Tidsskalaen er individenes alder.

For å illustrere forskjellen på enkel nøstet kasus-kontroll design og Counter-Matching har vi analysert data med begge metodene. I begge tilfeller er både Cox proporsjonal hasard modell og eksess risk modellen tilpasset. Datasettene som er brukt er i begge tilfeller 1:3, det vil si at tre kontroller er samlet pr kasus. Ved counter-matching ble individene delt i fire strata etter verdien av kumulativ radoneksponering. De fire gruppene ble definert ved kvartilene til kasusenes kumulative radoneksponering. Ved hvert hendelsestidspunkt  $t_j$  ble det trukket en kontroll fra hvert stratum, bortsett fra stratum til kasus. Til sammen utgjør de tre kontrollene og kasus den samlede risikomengden  $\tilde{\mathcal{R}}_j$  ved tid  $t_t$ .

Cox proporsjonal hasard modell har risikofunksjon gitt som (2.5), mens den for eksess relativ risiko modellen er

$$r(\beta, \mathbf{x}_i(t)) = [1 + x_{i1}(t)\beta_1][1 + x_{i2}(t)\beta_2], \quad (3.11)$$

der  $x_{i1}(t)$  og  $x_{i2}(t)$  er kumulativ radoneksponering og røykeeksponering henholdsvis, til tid  $t$ , for individ  $i$ .

Tabell 3.1: De estimerte regresjonskoeffisientene og standardavvik i eksempel 5.

Metode		Koef	St. avvik
Cox			
Kasus-kontroll	$\hat{\beta}_1$	0.0392	0.0055
	$\hat{\beta}_2$	0.0504	0.0092
Counter-Matching	$\hat{\beta}_1$	0.0424	0.0041
	$\hat{\beta}_2$	0.0587	0.0031
Eksess relativ risiko			
Kasus-Kontroll	$\hat{\beta}_1$	0.5561	0.2149
	$\hat{\beta}_2$	0.2760	0.0927
Counter-Matching	$\hat{\beta}_1$	0.4199	0.1364
	$\hat{\beta}_2$	0.2047	0.0678

Tabell 3.1 viser de estimerte regresjonskoeffisientene for både counter-matching og enkel nøstet kasus-kontroll. Som ventet blir standardavviket til koeffisientene mindre ved counter-matching enn ved enkel nøstet kasus-kontroll. Dette gjelder for begge modellene. I kapittel 5 vil vi studere forskjellen på de to samplingdesignene mer systematisk.

### 3.5 Telleprosessformulering av kasus-kontroll

Vi gjør i det følgende de samme antagelser som for kohortdata. Hendelsestidene  $t_j$  antas å ligge i intervallet  $(0, \tau]$ , der  $0 < \tau < \infty$ . Det er  $n$  individer i kohorten. Den merkede punktprosessen  $\{(t_j, i_j); j \geq 1\}$  registrerer for hver hendelsestid  $t_j$ , individet  $i_j$  som opplevde en hendelse ved tid  $t_j$ . Den individuelle telleprosessen

$$N_i(t) = \sum_{j \geq 1} I\{t_j \leq t, i_j = i\}$$

teller antall hendelser for individ  $i$  over  $(0, t]$ ,  $i = 1, 2, \dots, n$ . Den tilhørende intensitetsprosessen blir generelt

$$\lambda_i(t) = Y_i(t)\alpha_i(t)$$

der  $Y_i(t)$  som for kohort-data er en indikator som tar verdien 1 hvis individ  $i$  er under risiko ved tid  $t$  og 0 ellers, mens  $\alpha_i(t)$  er hasardraten til individet. Vi vil se på tilfellet med semiparametriske modeller hvor vi får intensitetsprosessen på formen

$$\lambda_i(t) = Y_i(t)r(\beta_0, \mathbf{x}_i(t))\alpha_0(t).$$

Så langt er alt analogt med kohortdata. Dette er på grunn av antagelsen om at individenes dødssannsynligheter ikke påvirkes av at vi samler kontroller. I det følgende presenteres verktøy som vil vise seg nyttige for å analysere de ulike samplingstrategiene, både ved hjelp av martingal residualprosesser og ved å brukes i mer formelle tester.

Vi introduserer  $\{(t_j, (i_j, \tilde{\mathcal{R}}_j)); j \geq 1\}$  som den merkede punktprosessen som for alle hendelsestidene i  $(0, t]$ , inneholder informasjon om hvilket individ som døde, samt risikomengdene som ble samplet ved hvert tidspunkt. Merkerommet  $E$  er endelig og gitt som

$$E = \{(i, \mathbf{r}) : i \in \{1, 2, \dots, n\}, \mathbf{r} \in \mathcal{P}_i\} = \{(i, \mathbf{r}) : \mathbf{r} \in \mathcal{P}_i, i \in \mathbf{r}\}.$$

For hver  $(i, \mathbf{r}) \in E$  har vi en telleprosess

$$N_{(i, \mathbf{r})}(t) = \sum_{j \geq 1} I\{t_j \leq t, (i_j, \tilde{\mathcal{R}}_j) = (i, \mathbf{r})\} \quad (3.12)$$

som teller antall ganger en hendelse inntreffer for individ  $i$  i intervallet  $(0, t]$ , der risikomengde  $\mathbf{r}$  ble samplet. Ved å aggregere over alle mulige risikomengder finnes den individuelle telleprosessen til individ  $i$

$$N_i(t) = \sum_{\mathbf{r} \in \mathcal{P}_i} N_{(i, \mathbf{r})}(t) = \sum_{j \geq 1} I\{t_j \leq t, i_j = i\} \quad (3.13)$$

som teller hendelser inntruffet for individ  $i$  over  $(0, t]$ , uansett hvilken risikomengde som har blitt samplet. Tilsvarende kan vi aggregere over alle individene og finne telleprosessene til hver risikomengde  $\mathbf{r}$

$$N_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} N_{(i, \mathbf{r})}(t) = \sum_{j \geq 1} I\{t_j \leq t, \tilde{\mathcal{R}}_j = \mathbf{r}\} \quad (3.14)$$

som teller antall ganger risikomengde  $\mathbf{r}$  har blitt samplet, uavhengig av hvilket individ som døde. Siden  $\lambda_{(i,\mathbf{r})}(t)dt = P(dN_{(i,\mathbf{r})}(t) = 1 | \mathcal{F}_{t-})$  er intensitetsprosessen til telleprosessen (3.12) gitt som

$$\lambda_{(i,\mathbf{r})}(t) = \lambda_i(t)\pi_t(\mathbf{r}|i).$$

Siden vi antar at den individuelle intensitetsprosessen kan uttrykkes som (2.4), og ved å benytte faktoriseringen (3.1) kan intensitetsprosessen skrives som

$$\lambda_{(i,\mathbf{r})}(t) = Y_i(t)r(\beta_0, \mathbf{x}_i(t))w_i(t, \mathbf{r})\pi_t(\mathbf{r})\alpha_0(t). \quad (3.15)$$

Intensitetsprosessen til (3.14) finnes ved å aggregere over (3.15). Intensitetsprosessen til prosessen som teller forekomsten av samplet risikomengde  $\mathbf{r}$  blir dermed

$$\lambda_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} Y_i(t)r(\beta_0, \mathbf{x}_i(t))w_i(t, \mathbf{r})\pi_t(\mathbf{r})\alpha_0(t).$$

Ved å se på forskjellen mellom telleprosessen og den kumulative intensitetsprosessen får vi den tilhørende martingalen til individ  $i$  med samplet risikomengde  $\mathbf{r}$

$$M_{(i,\mathbf{r})}(t) = N_{(i,\mathbf{r})}(t) - \Lambda_{(i,\mathbf{r})}(t) \quad (3.16)$$

der  $\Lambda_{(i,\mathbf{r})}(t)$  er den kumulative intensitetsprosessen til individ  $i$  med samplet risikomengde  $\mathbf{r}$ . Denne er gitt som

$$\Lambda_{(i,\mathbf{r})}(t, \beta_0) = \int_0^t Y_i(u)r(\beta_0, \mathbf{x}_i(u))w_i(u, \mathbf{r})\pi_u(\mathbf{r})\alpha_0(u)du. \quad (3.17)$$

Mens vi kan observere telleprosessen  $N_{(i,\mathbf{r})}(t)$  må vi estimere den kumulative intensitetsprosessen  $\Lambda_{(i,\mathbf{r})}(t)$ . Regresjonskoeffisientene  $\beta_0$  estimeres som vist i (3.4) med maximum partial likelihood estimatoren  $\hat{\beta}$ , mens vi estimerer  $\alpha_0(u)du$  med  $d\hat{A}_{0\mathbf{r}}(u, \hat{\beta})$ , der

$$\begin{aligned} \hat{A}_{0\mathbf{r}}(t, \beta_0) &= \sum_{t_j \leq t, \tilde{\mathbf{r}}_j = \mathbf{r}} \frac{1}{\sum_{l \in \mathbf{r}} Y_l(t_j)r(\beta_0, \mathbf{x}_l(t_j))w_l(t_j, \mathbf{r})\pi_{t_j}(\mathbf{r})} \\ &= \int_0^t \frac{dN_{\mathbf{r}}(s)}{\sum_{l \in \mathbf{r}} Y_l(s)r(\beta_0, x_l(s))w_l(s, \mathbf{r})\pi_s(\mathbf{r})Y_l(s)}. \end{aligned} \quad (3.18)$$

Vi setter (3.18) inn i (3.17) og får

$$\begin{aligned} \hat{\Lambda}_{(i,\mathbf{r})}(t, \beta_0) &= \int_0^t Y_i(s)r(\beta_0, \mathbf{x}_i(s))w_i(s, \mathbf{r})\pi_u(\mathbf{r})d\hat{A}_{0\mathbf{r}}(s) \\ &= \int_0^t \frac{Y_i(s)r(\beta_0, x_i(s))w_i(s, \mathbf{r})}{\sum_{l \in \mathbf{r}} Y_l(s)r(\beta_0, x_l(s))w_l(s, \mathbf{r})}dN_{\mathbf{r}}(s). \end{aligned} \quad (3.19)$$

Merk at det bare er baseline hasard som er estimert i (3.19). I risikofunksjonen inngår fortsatt  $\beta_0$ . Siden

$$\begin{aligned} N_{(i,\mathbf{r})}(t) - \hat{\Lambda}_{(i,\mathbf{r})}(t, \beta_0) &= \Lambda_{(i,\mathbf{r})}(t) + M_{(i,\mathbf{r})}(t) \\ &\quad - \int_0^t \frac{Y_i(s)r(\beta_0, x_i(s))w_i(s, \mathbf{r})}{\sum_{l \in \mathbf{r}} Y_l(s)r(\beta_0, x_l(s))w_l(s, \mathbf{r})}(dM_{\mathbf{r}}(s) + d\Lambda_{\mathbf{r}}(s)), \end{aligned}$$

der

$$d\Lambda_{\mathbf{r}}(s) = \sum_{l \in \mathbf{r}} Y_l(s) r(\beta_0, x_l(s)) w_l(s, \mathbf{r}) \pi_s(\mathbf{r}) \alpha_0(s) ds$$

får vi at

$$N_{(i, \mathbf{r})}(t) - \hat{\Lambda}_{(i, \mathbf{r})}(t, \beta_0) = M_{(i, \mathbf{r})}(t) - \int_0^t \frac{Y_i(s) r(\beta_0, x_i(s)) w_i(s, \mathbf{r})}{\sum_{l \in \mathbf{r}} Y_l(s) r(\beta_0, x_l(s)) w_l(s, \mathbf{r})} dM_{\mathbf{r}}(s). \quad (3.20)$$

Vi vet at integraler på formen  $I(t) = \int_0^t H(s) dM(s)$ , der  $H(s)$  er en stokastisk prosess hvis verdi i  $t$  er kjent rett før tid  $t$ , er et stokastisk integral som også er en martingal. Siden høyre side i (3.20) består av en martingal og et stokastisk integral og siden summen av to martingaler også er en martingal, har vi dermed vist at  $N_{(i, \mathbf{r})}(t) - \hat{\Lambda}_{(i, \mathbf{r})}(t, \beta_0) = \hat{M}_{(i, \mathbf{r})}(t, \beta_0)$  er en martingal. Martingalegenskapen beholdes dermed selv om vi estimerer baseline hasard.

For å estimere martingal residualprosessen må vi også estimere regresjonskoeffisientene til risikofunksjonen. Da blir differansen mellom telleprosessen og intensitetsprosessen bare tilnærmet en martingal. Vi finner den estimerte martingal residualprosessen som

$$\hat{M}_{(i, \mathbf{r})}(t) = N_{(i, \mathbf{r})}(t) - \hat{\Lambda}_{(i, \mathbf{r})}(t, \hat{\beta})$$

Martingalresidualene får vi når vi evaluerer martingal residualprosessen i tid  $t = \tau$ , der  $\tau$  er endetidspunktet for studien. Disse blir dermed gitt som

$$\hat{M}_{(i, \mathbf{r})} = \hat{M}_{(i, \mathbf{r})}(\tau) = N_{(i, \mathbf{r})}(\tau) - \hat{\Lambda}_{(i, \mathbf{r})}(\tau). \quad (3.21)$$

Martingalresidualene er forskjellen på observert og estimert antall hendelser for hver mulige kombinasjon  $(i, \mathbf{r})$ , og dermed et naturlig verktøy for å evaluere modeltilpasning.

De estimerte martingal residualprosessene vil vi bruke til å studere hvor godt ulike modeller passer til data. Generelt blir de estimerte martingal residualprosessene for kasus-kontroll

$$\hat{M}_{(i, \mathbf{r})}(t) = N_{(i, \mathbf{r})}(t) - \sum_{t_j \leq t} \frac{Y_i(t_j) r(\hat{\beta}, x_i(t_j)) w_i(t, \mathbf{r})}{\sum_{l \in \tilde{\mathbf{r}}_j} Y_l(t_j) r(\hat{\beta}, x_l(t_j)) w_l(t, \mathbf{r})},$$

der vektene bestemmes av kontrollsamplingen. Selv om  $\hat{M}_{(i, \mathbf{r})}(t)$  er definert for alle  $t \in (0, \tau]$  vil den være konstant etter at individ  $i$  har opplevd en hendelse. Dette fordi verken telleprosessen eller den kumulative intensitetsprosessene forandres etter dette. Etter hendelsestidspunktet  $t_j$  vil  $N_{(i, \mathbf{r})}(t)$  være konstant lik 1. Den kumulative intensitetsprosessen vil også holde seg konstant siden  $Y_i(t_j)$  er 0 for  $t > t_j$ .

Med stor kohort og liten samplet risikomengde blir det svært mange mulige samplete risikomengder  $\mathbf{r}$ . De fleste estimerte martingal residualprosessene  $\hat{M}_{(i, \mathbf{r})}(t)$  vil derfor bli 0. De individuelle residualprosessene til hvert par  $(i, \mathbf{r})$  er følgelig lite interessante. I neste avsnitt vil vi introdusere grupperte martingal residualprosesser. Ved å summere  $\hat{M}_{(i, \mathbf{r})}(t)$  for en gruppe individer etter en eller to kovariater vil summen av deres residualprosesser kunne gi oss innsikt i effekten av variablene vi har gruppert etter, samt fungere som et verktøy for å evaluere modellen som er brukt.

### 3.6 Grupperte martingal residualprosesser

For å studere kovariaters effekt på levetider ser vi gjerne på grupperte martingalresidualprosesser, gitt ved å aggregere over de individuelle martingal residualprosessene  $M_{(i,\mathbf{r})}(t)$  for en gruppe individer. De grupperte martingalresidualprosessene blir forskjellen mellom observert og estimert antall hendelser i hver gruppe. Individene deles inn i  $J = 1, \dots, G$  grupper, vanligvis basert på verdien til en eller to kovariater. Grupperingen kan avhenge av tid, men som vanlig må informasjonen som grupperingen ved tid  $t$  baseres på, være tilgjengelig rett før tid  $t$ , altså  $\mathcal{F}_{t-}$ . Den estimerte kumulative martingal residualprosessen til gruppe  $J$  ved tid  $t$  blir

$$\hat{M}_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i} d\hat{M}_{(i,\mathbf{r})}(u), \quad (3.22)$$

der  $\mathcal{J}(u)$  er mengden individer i gruppe  $J$  ved tid  $u$ . Ved å bruke (3.21) kan vi skrive høyre side av (3.22) som

$$\hat{M}_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i} dN_{(i,\mathbf{r})}(u) - \int_0^t \sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i} d\hat{\Lambda}_{(i,\mathbf{r})}(u). \quad (3.23)$$

Ved å summere telleprosessene over alle mulige samlete risikomengder finner vi de individuelle telleprosessene, jmf (3.13). Videre kan inkrementene til den kumulative intensitetsprosessen skrives som (3.19) og dermed får vi at den estimerte martingal residualprosessen til gruppe  $J$  kan uttrykkes som

$$\hat{M}_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} dN_i(u) - \int_0^t \sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i} \frac{Y_i(s)r(\beta, x_i(s))w_i(s, \mathbf{r})}{\sum_{l \in \mathbf{r}} Y_l(s)r(\beta, x_l(s))w_l(s, \mathbf{r})} dN_{\mathbf{r}}(s). \quad (3.24)$$

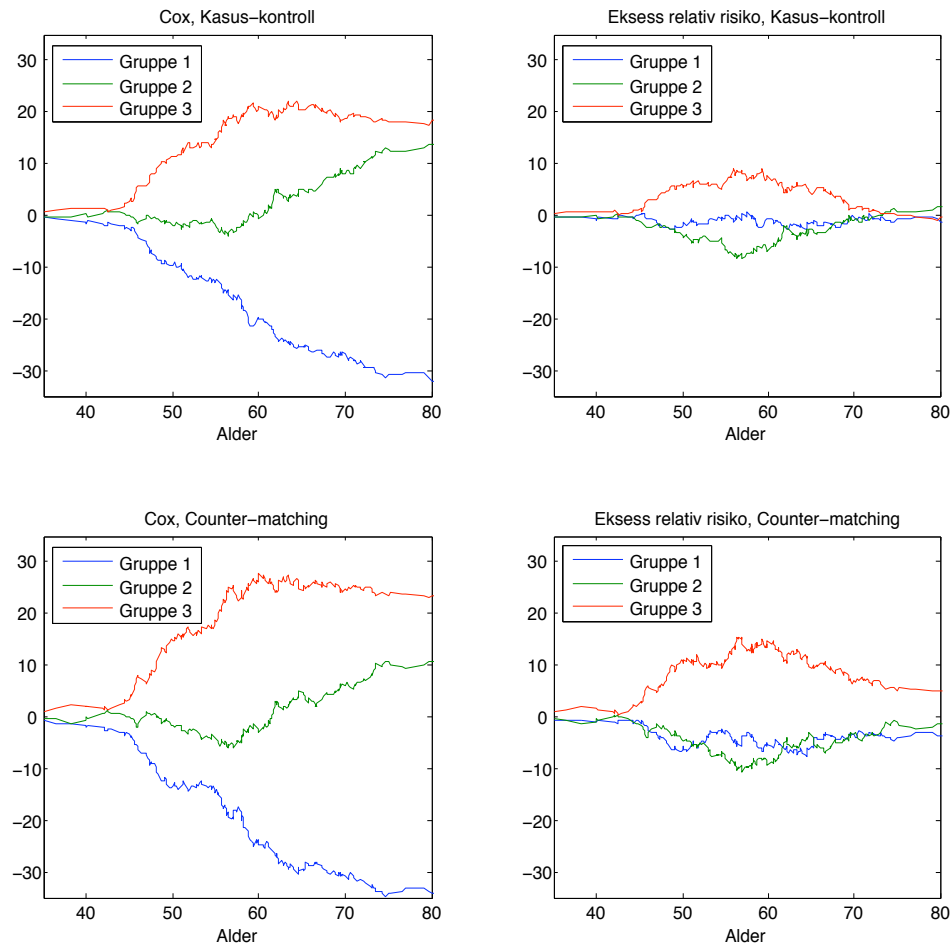
Vi kan alltid bytte om på summerekkefølgen. Vi kan derfor skrive  $\sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i}$  som  $\sum_{\mathbf{r} \in \mathcal{P}} \sum_{i \in \mathbf{r} \cap \mathcal{J}(u)}$ . Siden  $dN_{\mathbf{r}}(s)$  hopper med én hver gang en risikomengde  $\mathbf{r}$  samples, som er ved hver hendelsestid kan vi skrive (3.24) som

$$\hat{M}_J(t) = N_J(t) - \sum_{t_j \leq t} \frac{\sum_{i \in \tilde{\mathcal{R}}_j \cap \mathcal{J}(t_j)} r(\hat{\beta}, x_i(t_j))w_i(t_j, \tilde{\mathcal{R}}_j)}{\sum_{l \in \tilde{\mathcal{R}}_j} r(\hat{\beta}, x_l(t_j))w_l(t_j, \tilde{\mathcal{R}}_j)}, \quad (3.25)$$

der  $N_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} dN_i(u)$ . Vektene  $w_i(t_j, \tilde{\mathcal{R}}_j)$  bestemmes av designet for kontroll-samplingen. Ved counter-matching blir dermed den estimerte martingal residualprosessen som (3.25) der vektene er gitt i (3.3). Ved enkel nøstet kasus kontroll blir vektene (3.2) like for alle individer og dermed forkortet. Den estimerte martingal residualprosessen til gruppe  $J$  blir for kasus-kontroll dermed

$$\hat{M}_J(t) = N_J(t) - \sum_{t_j \leq t} \frac{\sum_{i \in \tilde{\mathcal{R}}_j \cap \mathcal{J}(t_j)} r(\hat{\beta}, x_i(t_j))}{\sum_{l \in \tilde{\mathcal{R}}_j} r(\hat{\beta}, x_l(t_j))}.$$

Merk at som likelihooden er den estimerte martingal residualprosessen ved enkel nøstet kasus kontroll et spesialtilfelle av counter-matching, der kohorten bare deles inn i ett stratum.



Figur 3.1: Martingal residualprosessene i eksempel 6. Tilpasset modell og samplingdesign er gitt i overskriften til hvert plot. Gruppering etter kumulativ radoneksponering. Gruppe 1: 0 – 500 WLM, Gruppe 2: 500 – 1500 WLM, Gruppe 3: Over 1500 WLM.

**Eksempel 6 (Grupperte martingal residualprosesser.)** For å evaluere modellene tilpasset på levetidsdataene i de foregående eksemplene, ble de estimerte martingal residualprosessene beregnet. Individene ble delt i tre grupper etter verdien av kumulativ radon eksponering. Gruppe 1 er alle består av alle individer med kumulativ radoneksponering under 500 WLM. Gruppe 2 består av alle med kumulativ radoneksponering mellom 500 og 1500 WLM, mens gruppe 3 utgjøres av individene med kumulativ radoneksponering over 1500 WLM.

Figur 3.1 viser de estimerte martingal residualprosessene for Cox-modellen og eksess relativ risiko modellen, både for enkel nøstet kasus-kontroll og counter-matching. Plottet av de estimerte martingal residualprosessene til Cox-modellen med enkel nøstet kasus-kontroll, viser at det er store avvik mellom observerte og forventede antall dødsfall. For gruppe 1 som er individene med lavest eksponering for radon predikerer modellen flere dødsfall

enn det som observeres for alle aldre. For gruppe 2 passer modellen bra inntil omkring 60 år, for deretter å predikere færre dødsfall enn det som observeres. For gruppe 3 som er individene med høyest verdi av radoneksponering undervurderer modellen dødeligheten fram til rundt 60 år, for deretter å predikere noe færre dødsfall enn det som observeres.

Plottet av martingal residualprosessen til Cox-modellen med 1 : 3 counter-matchet risikomengde viser at avvikene fra modellen blir noe større enn ved enkel nøstet kasus kontroll. Forskjellen mellom gruppene er ellers den samme for begge samplingmetodene.

Dersom effekten av kovariater endres over tid, vil en modell med konstante effekter av kovariatene tilpasse et middelnivå av effekten, som vil være for lavt ved noen tidspunkter og for høyt ved andre. For en kovariat med høy effekt i lav alder og lavere effekt ved høyere alder, vil en modellen estimere en konstant effekt av kovariaten som vil være for lav ved lav alder og for høy senere. Individer med høye verdier av denne variabelen vil da ha høyere dødelighet i ung alder enn modellen predikerer, og lavere dødelighet senere i livet. For individer med lavere verdier av kovariaten vil vi se det motsatte. Modellen predikerer for høy dødelighet til å begynne med og for lav senere.

Disse effektene ser vi i plottet av martingal residual prosessen for eksess relativ risiko modellen. I tilfellet med enkel nøstet kasus-kontroll ser vi at spesielt gruppe 2 har høyere dødelighet enn modellen predikerer inntil ca. 55 år, med stor negativ residual som resultat. Etter 55 år predikerer modellen flere dødsfall enn det som observeres og residualen ender rundt 0. For gruppe 3 har residualen akkurat den motsatte utviklingen. Gruppe 1 ser derimot modellen ut til å passe godt for.

Ved 1 : 3 counter-matching ser vi at residualene til eksess relativ risiko modellen blir større enn med enkel nøstet kasus-kontroll. Plottet viser også at den predikerte dødeligheten til gruppe 1 som for kasus-kontroll stemte bra med det observerte, nå blir undervurdert fram til rundt 60 år, for deretter å bli noe overvurdert.

Konklusjonen basert på plot av de grupperte estimerte martingal residualprosessene blir at de store avvikene for alle modellene tyder på at ingen av modellene passer godt til data. Cox-modellen ser også ut til å passe enda dårligere enn eksess relativ risiko modellen. I neste avsnitt vil vi introdusere en kji-kvadrat goodness-of-fit test som en annen metode for å sjekke modelltilpasningen.

I neste avsnitt har vi bruk for estimator for kovariansen mellom de grupperte martingal residualprosessene. Vi kan vise at kovariansen mellom  $\hat{M}_I(s)$  og  $\hat{M}_J(t)$  kan estimeres ved

$$\hat{\sigma}_{IJ}(s, t) = \phi_{IJ}(0, s \wedge t, \hat{\beta}) - \hat{\psi}_I(0, s, \hat{\beta})^T I(\hat{\beta})^{-1} \hat{\psi}_J(0, t, \hat{\beta}),$$

der

$$\phi_{IJ}(s_1, s_2, \hat{\beta}) = \sum_{s_1 \leq t_j \leq s_2} \frac{S_{\tilde{R}_j I}^{(0)}(\hat{\beta}, t_j)}{S_{\tilde{R}_j}^{(0)}(\hat{\beta}, t_j)} \left\{ \delta_{IJ} - \frac{S_{\tilde{R}_j J}^{(0)}(\hat{\beta}, t_j)}{S_{\tilde{R}_j}^{(0)}(\hat{\beta}, t_j)} \right\}.$$

Her er  $\delta_{IJ}$  Kronecker delta som er definert 1 hvis  $I = J$  og 0 hvis  $I \neq J$ , og

$$\hat{\psi}_J(s_1, s_2, \hat{\beta}) = \sum_{s_1 \leq t_j \leq s_2} \left\{ \frac{S_{\tilde{R}_j J}^{(1)}(\hat{\beta}, t_j)}{S_{\tilde{R}_j}^{(0)}(\hat{\beta}, t_j)} - \frac{S_{\tilde{R}_j J}^{(0)}(\hat{\beta}, t_j) S_{\tilde{R}_j J}^{(1)}(\hat{\beta}, t_j)}{S_{\tilde{R}_j}^{(0)}(\hat{\beta}, t_j)^2} \right\}$$

der  $S_{\tilde{R}_j}^{(0)}(\beta, t_j)$  og  $S_{\tilde{R}_j}^{(1)}(\beta, t_j)$  er gitt ved hhv (3.6) og (3.7), mens

$$\begin{aligned} S_{\tilde{R}_j J}^{(0)}(\beta, t_j) &= \sum_{l \in \tilde{R}_j \cap J(t_j)} r(\beta, x_l(t_j)) w_l(t_j, \tilde{R}_j) \\ S_{\tilde{R}_j J}^{(1)}(\beta, t_j) &= \sum_{l \in \tilde{R}_j \cap J(t_j)} \dot{r}(\beta, x_l(t_j)) w_l(t_j, \tilde{R}_j) \end{aligned}$$

### 3.7 Kji-kvadrat goodness-of-fit test

Hittil har vi sett hvordan plot av martingal residualprosesser kan brukes til å sjekke modelltilpasningen i overlevelsesanalysen. Mens dette er metoder som baserer seg på skjønn, vil vi nå se på en mer formell måte å kontrollere modelltilpasningen på. Vi er interessert i å teste om vi har tilpasset riktig modell. Vi vil introdusere en enkel kji-kvadrat goodness-of-fit test basert på sammenligning av observert og forventet antall hendelser i de  $G$  gruppene i  $K$  disjunkte tidsintervaller.

Vi deler tidsintervallet inn i  $K$  grupper slik at  $0 < a_0 < a_1 < \dots < a_{K-1} < a_K < \tau$ , der  $0 < \tau < \infty$ . Da blir forskjellen mellom observert og estimert antall hendelser i gruppe  $J$  i tidsintervall  $H$

$$\hat{M}_{HJ} = \hat{M}_J(a_H) - \hat{M}_J(a_{H-1}) = O_{HJ} - E_{HJ},$$

for  $H = 1, 2, \dots, K, J = 1, 2, \dots, G$ , og  $O_{HJ} = N_J(a_H) - N_J(a_{H-1})$ , mens

$$E_{HJ} = \sum_{a_{H-1} < t_j < a_H} \frac{\sum_{i \in \tilde{R}_j \cap J(t_j)} r(\hat{\beta}, x_i(t_j)) w_i(t_j, \tilde{R}_j)}{\sum_{i \in \tilde{R}_j} r(\hat{\beta}, x_i(t_j)) w_i(t_j, \tilde{R}_j)} \quad (3.26)$$

er antall forventede observasjoner under den semiparametriske modellen. Siden martingal residualprosessene summerer til 0 ved hver tid  $t$  ser vi bort fra gruppe 1, og ser på  $K(G-1)$ -vektoren  $\hat{M}$  med elementer  $\hat{M}_{HJ}$  for  $H = 1, 2, \dots, K, J = 2, \dots, G$ . Videre blir  $\hat{M}$  tilnærmet multinormalfordelt med forventning 0 i store utvalg når modellen stemmer.

Kovariansmatrisen kan estimeres ved matrisen  $\hat{\Sigma} = \{\hat{\sigma}_{LI, HJ}\}$ , med elementer

$$\begin{aligned} \hat{\sigma}_{LI, HJ} &= \text{Cov}(\hat{M}_{LI}, \hat{M}_{HJ}) \\ &= \delta_{LH} \phi_{IJ}(a_{H-1}, a_H, \hat{\beta}) - \hat{\psi}_I(a_{L-1}, a_L, \hat{\beta})^T I(\hat{\beta})^{-1} \hat{\psi}_J(a_{H-1}, a_H, \hat{\beta})^T, \end{aligned}$$

for  $H, L = 1, 2, \dots, K, J = 2, 3, \dots, G$ .

Goodness-of-fit kji-kvadrat testen kan derfor baseres på observatoren

$$\chi^2 = \hat{M}^T \hat{\Sigma}^{-1} \hat{M} \quad (3.27)$$



som i store utvalg er tilnærmet kji-kvadrat fordelt med  $K(G-1)$  frihetsgrader når modellen stemmer.

**Eksempel 7 (Kji-kvadrat goodness-of-fit test.)** For videre å studere forskjellen på de to samplingdesignene, er testen (3.27) brukt på lungekreftdataene presentert i de foregående eksemplene. For counter-matching ble individene delt i de samme tre gruppene etter radoneksponering som i eksempel 6. De ble også gruppert etter alder, etter hvorvidt de er over eller under 60 år. Både for enkel nøstet kasus-kontroll og counter-matching, ble det som tidligere benyttet 1:3 sampling av kontroller.

Tabell 3.2: Observert og estimert antall dødsfall for lungekreftdataene, for counter-matching (CM) og enkel nøstet kasus-kontroll (ENKK).

Gruppe	Observert	Estimert ENKK	Estimert CM
Eksess relativ risiko			
Under 60 år			
Gruppe 1	30	30.7	35.5
Gruppe 2	39	45.9	48.4
Gruppe 3	81	73.4	66.1
Over 60 år			
Gruppe 1	27	27.7	25.3
Gruppe 2	45	36.1	36.9
Gruppe 2	36	44.2	45.8
Cox			
Under 60 år			
Gruppe 1	30	50.2	54.7
Gruppe 2	39	39.8	42.0
Gruppe 3	81	60.0	53.3
Over 60 år			
Gruppe 1	27	39.2	36.6
Gruppe 2	45	29.9	31.2
Gruppe 2	36	38.9	40.2

Tabell 3.2 bekrefter konklusjonen fra eksempel 6. Cox-modellen predikerer for mange dødsfall for individer med lave verdier av radoneksponering, mens den predikerer for få dødsfall for individene med høyest verdier. Dette gjelder uavhengig av samplingmetode, men er langt klarere for aldersgruppen under 60 år enn for aldersgruppen over 60 år. I alle bortsett fra et tilfelle er avvikene større for counter-matching enn enkel nøstet kasus-kontroll. For individene i den midterste gruppen ser Cox-modellen ut til å passe ganske bra for individene under 60 år, mens modellen predikerer for få dødsfall for aldersgruppen over 60 år. Dette stemmer godt med figur 3.1 som viser at den estimerte martingal residualprosessen for Cox-modellen, uansett samplingdesign, ligger rundt 0 inntil ca. 60, for deretter å stige.

For eksess relativ risiko modellen er avvikene mindre enn for Cox-modellen. Fortsatt ligger

estimatene basert på enkel nøstet kasus-kontroll nærmest det observerte i de fleste tilfellene, men forskjellene er mindre enn ved Cox-modellen. Resultatene er som forventet i forhold til figur 6. Modellen passer stort sett bra for gruppe 1 uansett alder. Unntaket er for counter-matching under 60 år der modellen spår for mange dødsfall. For gruppe 2 predikerer modellen for mange dødsfall før 60 og for få dødsfall etter 60 år. Det motsatte gjelder for gruppe 3.

Når eksess relativ risiko modellen tilpasses blir  $\chi^2$ -kvadrat goodness-of-fit observatoren 10.5 for enkel nøstet kasus-kontroll, mens den for counter-matching blir 14.2. Sammenlignet med  $\chi^2$ -kvadrat fordelingen med  $2(3 - 1) = 4$  frihetsgrader blir p-verdiene 0.032 og 0.007 henholdsvis. Når Cox-modellen benyttes blir resultatene 69.5 og 75.8 for enkel nøstet kasus-kontroll og counter-matching henholdsvis. De tilhørende p-verdiene blir mindre enn 0.001. Konklusjonene fra eksempel 6 styrkes dermed. Ingen av modellene passer godt til data, men eksess relativ risiko-modellen passer langt bedre enn Cox-modellen.

I dette kapitlet har vi introdusert analyseverktøy for studier med samlet risikomengde. Vi har sett hvordan vi kan tilpasse en regresjonsmodell og hvordan vi kan evaluere tilpassningen av modellen ved hjelp av plot av grupperte estimerte martingal residual prosesser samt goodness-of-fit testen (3.27). Senere i oppgaven vil vi studere egenskapene til denne goodness-of-fit testen ved hjelp av simuleringer.

## Kapittel 4

# Om simuleringene

### 4.1 Innledning

Hittil har vi sett på metoder for å analysere overlevelseshdata. I kapittel 5 vil vi bruke metodene presentert i kapittel 2 og 3 til å analysere simulerte overlevelseshdata. Hensikten med simuleringene vil være å undersøke signifikansnivået og styrken til goodness-of-fit testen (3.27) i ulike situasjoner.

I dette kapitlet vil vi gjøre rede for hvordan simuleringene er gjennomført og hvordan utfordringer som har dukket opp underveis er løst. I avsnitt 4.2 vil vi presentere hvordan levetider simuleres og hvordan vi kan oppnå ønsket andel sensurerte levetider. I avsnitt 4.3 vil vi se hvordan vi kan utvide simuleringsmetoden til å kunne simulere levetider der en kovariat er tidsavhengig, eller transformert. I avsnitt 4.4 vil vi se nærmere på ulike problemer som dukket opp i simuleringene og hvordan disse ble løst.

### 4.2 Simulering av levetider

For å gjøre simuleringene våre har vi behov for å trekke levetider fra en fordeling med spesifisert kumulativ hasard  $H(t)$ . Ved å la levetiden  $T$  være gitt ved  $T = H^{-1}(V)$ , der  $V$  er eksponential fordelt med parameter 1, får vi overlevelseshfunksjon

$$\begin{aligned} S(t) &= P(T > t) \\ &= P(H^{-1}(V) > t) \\ &= P(V > H(t)) \\ &= e^{-H(t)} \end{aligned}$$

Fra (2.2) har vi at overlevelseshfunksjonen kan skrives som  $S(t) = e^{-H(t)}$ , der  $H(t)$  er kumulativ hasard. Altså får levetiden nå kumulativ hasard  $H(t)$ . Det velges en form på  $H(t)$  og dermed  $H^{-1}(t)$ , slik at data får den formen vi ønsker. Merk at  $H(t)$  må være invertibel.

I denne oppgaven vil vi bruke semiparametriske regresjonsmodeller. Kumulativ hasard avhenger da av kovariatene  $X$  og vi skriver  $H_X(t)$ . For semiparametriske regresjonsmodeller er kumulativ hasard gitt som

$$H_X(t) = H(t|X) = H_0(t)r(\beta, X),$$

der  $r(\beta, X)$  er risikofunksjonen som bestemmes av valgt regresjonsmodell, mens  $H_0(t)$  er kumulativ baseline hasard. Vi har vist at vi kan trekke levetider med  $H_X(t)$  som kumulativ hasard ved å la  $T$  være gitt ved  $T = H_X^{-1}(V)$ , der  $V$  er eksponentialfordelt med parameter 1. Gitt at vi velger en strengt voksende, kontinuerlig funksjon  $H_0(t)$  for kumulativ baseline hasard kan vi finne et uttrykk for den inverse kumulative hasarden:

$$\begin{aligned} H_X(t) &= H_0(t) r(\beta, X) \\ H_X^{-1}(v) &= H_0^{-1}\left(\frac{v}{r(\beta, X)}\right) \end{aligned}$$

Levetiden  $T$  får altså ønsket fordeling ved å trekke den ved  $T = H_0^{-1}\left(\frac{V}{r(\beta, X)}\right)$ , der  $V \sim \exp(1)$ .

For å spesifisere den kumulative hasarden må det i tillegg til risikofunksjon velges form på kumulativ baseline hasard. Vi bruker her Weibull-fordeling og får dermed baseline hasard  $h_0(t) = a k t^{k-1}$ , mens den kumulative baseline hasarden blir  $H_0(t) = a t^k$ . Den inverse kumulative baseline hasarden  $H_0^{-1}(y)$  blir nå  $\sqrt[k]{\frac{y}{a}}$ , der  $k$  og  $a$  er kjente konstanter. Disse velges slik at de simulerte levetidene får ønsket fordeling. Mens  $a$  er en skalaparameter, vil forskjellige verdier av  $k$  produsere fordelinger der lange levetider reduseres mer og mer for høyere verdier av  $k$ . Dermed endres formen på levetidsfordelingen og vi sier at  $k$  er en formparameter. Verdier av  $k$  mindre enn 1 vil gi avtagende baseline hasard,  $k = 1$  gir konstant baseline, mens  $k > 1$  svarer til stigende baseline hasard. Den kumulative hasarden  $H_X(t)$  er nå fullstendig definert. Dermed kan vi trekke levetider  $T$ , som

$$T = H_X^{-1}(V) = H_0^{-1}\left(\frac{V}{r(\beta, X)}\right) = \sqrt[k]{\frac{V}{a r(\beta, X)}} \quad (4.1)$$

der  $V \sim \exp(1)$ . I denne oppgaven vil vi konsentrere oss om de to velkjente risikofunksjoner Cox semiparametriske regresjonsmodell (2.5) og eksess relativ risiko modellen (3.11).

Siden sensurering finner sted i de aller fleste virkelige studier er det viktig at vi også får det med i simuleringene. Til hver simulerte levetid  $T$  trekkes derfor en sensureringstid  $C$  fra en fordeling med tetthet  $g_C(t)$ . Til slutt finner vi den sensurerte levetiden  $\tilde{T}$  som  $\tilde{T} = \min(T, C)$ . For videre analyser må også en sensurindikator  $d = I(\tilde{T} = T)$  legges til datasettet. Denne er 1 dersom individet har dødd, 0 hvis det er sensurert.

Ved å variere parametrene i  $g_C(t)$  kan vi få ønsket andel sensurerte levetider. I neste avsnitt vil vi se hvordan vi kan beregne sannsynligheten for at en levetid sensureres, og dermed den forventede andelen sensurerte levetider. I eksempel 8 vil vi se hvordan resultatene fra avsnitt 4.2.1 kan anvendes i praksis.

### 4.2.1 Om sensurering i simuleringene

Ovenfor så vi hvordan sensurerte levetider  $\tilde{T}$  kan simuleres ved å først å trekke en levetid  $T$  og en sensureringstid  $C$ , for til slutt å velge den sensurerte levetiden  $\tilde{T}$  som den minste av  $T$  og  $C$ . Andelen tilfeller der  $C < T$  avhenger av fordelingen til  $C$ ,  $g_C(t)$ . Siden vi blant annet ønsker å studere hvordan andelen sensurerte levetider påvirker analysen vår, er det praktisk å kunne tune  $g_C(t)$  slik at ønsket andel sensureres. Dette gjøres ved å variere parameterene til  $g_C(t)$ . For å finne den verdien av parameterene som gir ønsket andel sensurering finner vi et uttrykk for  $P(C < T)$  som en funksjon av parameterene i  $g_C(t)$ . Dette finner vi som

$$\begin{aligned} P(C < T) &= E[I(C < T)] \\ &= E_X[E[I(C < T)|X]] \\ &= E_X[P(C < T|X)] \\ &= E_X \int_0^\infty P(T > u, C > u|X) h_C(u) du, \end{aligned} \quad (4.2)$$

der  $h_C(u)$  er hasarden for sensureringstiden  $C$ . Denne angir sannsynligheten for å sensureres i neste korte tidsintervall, gitt at man hittil ikke er sensurert. Dermed blir  $h_C(u) = \frac{g_C(u)}{1-G_C(u)}$ , der  $G_C(u)$  er den kumulative fordelingen til sensureringstiden  $C$ . Siden  $P(T > u|X) = e^{-H_X(u)}$ , og  $P(C > u) = 1 - G_C(u)$ , og siden  $G_C(u)$  er sannsynligheten for å bli sensurert en gang før  $u$ , kan vi skrive (4.2) som

$$\begin{aligned} P(C < T) &= E_X \int_0^\infty e^{-H_X(u)} (1 - G_C(u)) \frac{g_C(u)}{1 - G_C(u)} du \\ &= E_X \int_0^\infty e^{-H_X(u)} g_C(u) du. \end{aligned} \quad (4.3)$$

I denne oppgaven er sensureringstidene trukket fra en Weibull-fordeling. Weibull-fordelingen har tetthet  $g_C(t) = c f t^{f-1} e^{-c t^f}$ , der  $f$  og  $c$  er henholdsvis form og skalaparameter. Ved å sette denne inn i (4.3) kan sannsynligheten for å sensureres skrives som

$$P(C < T) = E_X \int_0^\infty e^{-H_X(u)} c f u^{f-1} e^{-c u^f} du. \quad (4.4)$$

Andelen som faktisk sensureres vil variere mellom de simulerte datasettene. For hvert enkelt individ er det en binomisk sannsynlighet  $p$  for at han vil sensureres som kan finnes med (4.4). Andelen sensurerte levetider i et simulert datasett  $\hat{p}$  blir da for stor nok  $n$  og  $p$  tilnærmet normalfordelt med forventning  $p$  og standardavvik  $\sqrt{\frac{p(1-p)}{n}}$ .

I våre simuleringer vil vi alltid trekke  $n = 1000$  levetider, mens vi ønsker å variere andelen sensurerte levetider mellom 80%, 90% og 95%. Standardavviket til andelen sensurerte levetider blir i de tre situasjonene henholdsvis 0.013, 0.009 og 0.007. Dette betyr for eksempel at når vi trekker sensureringstider fra en fordeling som er tunet slik at den forventede andelen sensurerte levetider er 0.95, vil den observerte andelen sensurerte levetider i omtrent 95% av simuleringene ligge i intervallet (0.936, 0.964).

**Eksempel 8 (Simuleringseksempel)** For å illustrere hvordan simulering og sensurering fungerer i praksis, vil vi gå gjennom et kort eksempel der vi bruker teorien presentert i de foregående avsnittene. Vi vil trekke 1000 levetider fra en spesifisert fordeling, med Cox risikofunksjon, sensurere levetidene slik at ønsket andel sensureres og deretter tilpasse Cox semiparametriske regresjonsmodell og se hvordan de estimerte regresjonskoeffisientene blir i forhold til de som ble brukt i simuleringene. Deretter vil vi utføre testen (3.27) på den estimerte modellen og se hvordan fordelingen til testobservatorene blir. Under nullhypotesen skal testobservatorene være  $\chi^2$ -kvadrat fordelt. Siden det er den sanne modellen som er tilpasset håper vi at ca  $100 \cdot \alpha\%$  av testene gir forkastning på signifikansnivå  $\alpha$ .

Vi begynner med å trekke to kovariater  $X_1, X_2$ , som uniform og diktom henholdsvis. For å få en enkel tolkning av baseline hasard og regresjonskoeffisientene, bestemmer vi at begge kovariatene skal ha minste verdi lik 0. På denne måten reduserer vi også muligheten for negativ estimert risikofunksjon når vi senere skal tilpasse eksess relativ risiko modellen til slike datasett. Det vil bli drøftet nøyere i avsnitt 4.4.1. For både Cox og eksess relativ risiko modellene vil et individ som har begge kovariatene lik 0 få risikofunksjon lik 1. For begge modellene får baseline hasard dermed tolkningen hasarden til et individ med minste mulige verdi av begge kovariatene.

Når regresjonskoeffisientene skal estimeres er det ønskelig at de to variablene har lik varians. Dette fordi det er lettere å estimere effekten av en variabel med stor variasjon, enn en variabel med mindre variasjon. Ved å gi de to variablene lik varians får vi dermed estimert effekten av dem med like stor nøyaktighet. Når vi har bestemt oss for utfallsrommet til den ene variabelen, velges den andre slik at de får lik varians. Ved å la  $X_1$  være  $\sim U(0, 1)$ , finner vi at den diktome variabelen  $X_2$  må ta verdiene 0 eller  $\frac{1}{\sqrt{3}}$  med samme sannsynlighet, for at de to variablene skal få lik varians og ha minste verdi lik 0. Ved å la de to variablene  $X_1, X_2$  være definert som

$$\begin{aligned} X_1 &\sim U(0, 1) \\ X_2 &= \begin{cases} 0, & \text{med sannsynlighet } \frac{1}{2} \\ \frac{1}{\sqrt{3}}, & \text{med sannsynlighet } \frac{1}{2}, \end{cases} \end{aligned}$$

vil altså ønsket om lik varians for de to variablene blitt oppfylt.

Som vist over finner vi nå levetiden  $T$  ved  $T = \sqrt[k]{\frac{V}{akr(\beta, X)}}$ , der  $V \sim \exp(1)$ , mens  $r(\beta, X)$  er risikofunksjonen. Vi velger å bruke Cox semiparametriske regresjonsmodell med risikofunksjon  $r(\beta, X) = e^{\beta^T X}$ , og levetidene  $T$  er da gitt ved  $T = \sqrt[k]{\frac{V}{ake^{\beta^T X}}}$ . Konstantene  $k$  og  $a$  settes i dette eksemplet til hhv. 2 og  $1.0 \cdot 10^{-2}$  slik at vi får passelig skalerte levetider. Regresjonskoeffisientene  $\beta = (\beta_1, \beta_2)^T$  settes begge til 1. På denne måten blir risikofunksjonen til et individ med maksimale verdier av kovariatene omtrent fem ganger så stor som for et individ med minimale verdier av kovariatene, når Cox-modellen brukes. Dette er et realistisk størrelsesforhold i forhold til virkelige medisinske studier.

Til slutt trekkes en sensureringstid  $C$  fra  $g_C(t)$ , der formparameteren  $f$  settes lik 1, mens  $c$  velges slik at ønsket andel sensureres. Dermed blir  $g_C(t) = ce^{-ct}$ . Ved å ta forventningen

mhp.  $X_1$  og  $X_2$  i (4.4) finner vi sannsynligheten for å sensureres som

$$\begin{aligned} P(C < T) &= E_X \int_0^\infty e^{-H(u|X)} ce^{-cu} du \\ &= \int_0^1 \int_0^\infty e^{-H(u|x_1, X_2=0)} f(x_1) P(X_2 = 0) ce^{-cu} du dx_1 \\ &\quad + \int_0^1 \int_0^\infty e^{-H(u|x_1, X_2=\frac{1}{\sqrt{3}})} f(x_1) P(X_2 = \frac{1}{\sqrt{3}}) ce^{-cu} du dx_1 \end{aligned} \quad (4.5)$$

der  $f(x_1)$  er sannsynlighetstettheten til  $X_1$ . Nå er  $f(x_1) = 1$ , mens  $P(X_2 = 0) = P(X_2 = \frac{1}{\sqrt{3}}) = \frac{1}{2}$ . Dermed kan vi skrive (4.5) som

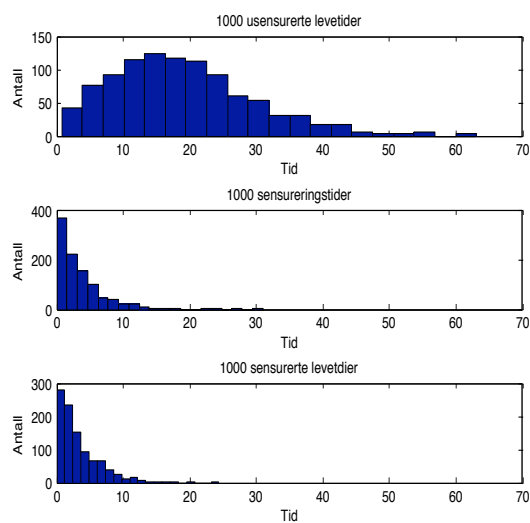
$$\begin{aligned} P(C < T) &= \frac{1}{2} \int_0^1 \int_0^\infty e^{-H(u|x_1, X_2=0)} ce^{-cu} du dx_1 \\ &\quad + \frac{1}{2} \int_0^1 \int_0^\infty e^{-H(u|x_1, X_2=\frac{1}{\sqrt{3}})} ce^{-cu} du dx_1. \end{aligned} \quad (4.6)$$

Vi kan nå velge  $c$  slik at ønsket andel sensureres. I praksis må integralet (4.6) løses numerisk. Integralet ble beregnet av matlabs funksjon for dobbeltintegraler, `dblquad`. Funksjonen for å finne nullpunkter i samme program, `fzero`, ble brukt til å løse ligningen  $P(C < T) = p$ , der  $p$  er ønsket andel sensurering, med hensyn på  $c$ . I dette eksemplet ønsker vi at 95% skal sensureres, og dermed må  $c$  være 0.2844.

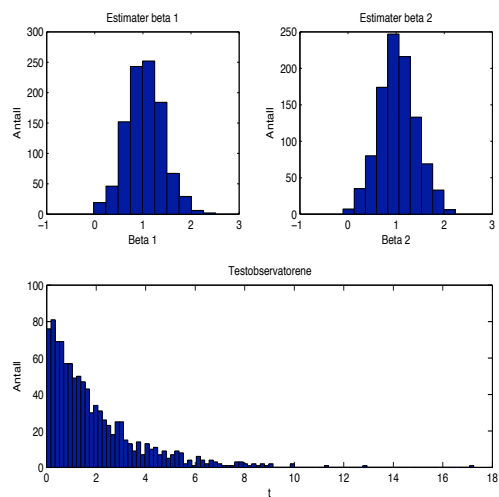
Metodene for å analysere data presentert i det forrige kapitelet baserer seg på å sammenligne kovariatinformasjonen til kasus med kovariatinformasjonen til kontrollene ved hvert hendelsestidspunkt. For hver simulerte hendelsestid må det derfor trekkes kontroller etter ønsket samplingdesign. Her er 1 : 1 enkel nøstet kasus-kontroll brukt. Det er med andre ord, ved hver hendelsestid  $t$ , kun trukket én kontroll tilfeldig blant individene som var under risiko ved tid  $t$ . Figur 4.1 viser hvordan fordelingen til de simulerte usensurerte levetidene, sensureringstidene og sensurerte levetidene ser ut. De estimerte regresjonskoeffisientene ble i gjennomsnitt 1.0394 og 1.0342. Som vi husker var de sanne verdiene 1 for begge.

Testen (3.27) baserer seg på å analysere forskjellen mellom observert og estimert antall hendelser inntruffet i spesifiserte intervaller. I dette eksemplet ble det gruppert etter tid og verdi av  $X_1$ . Tidsgruppene var over og under 6 som er i nærheten av medianen for de sensurerte levetidene, mens det etter kovariaten ble gruppert etter hvorvidt verdien av  $X_1$  var større eller mindre enn 0.5. Dermed ble det to tidgrupper og to kovariatgrupper. Antall frihetsgrader i testen ble da  $2 \cdot (2 - 1) = 2$ .

Det nederste plottet i figur 4.2 viser hvordan de 1000 testobservatorene ble fordelt. Gjennomsnittsverdien til testobservatorene ble 1.9265, mens forventningen er 2. Videre ga 4.7% av testene forkastning på signifikansnivå 5%. En bedre sjekk av hvorvidt testobservatorene har riktig fordeling, er det å lage et QQ-plot av de observerte kvantilene mot de teoretiske kvantilene fra  $\chi^2$ -fordelingen. Dette er gjort i figur 4.3. Figuren viser ingen tegn på at testobservatorene avviker fra  $\chi^2$ -fordelingen med 2 frihetsgrader. I samtlige av de 1000 kjøringene konvergente scoring-algoritmen.

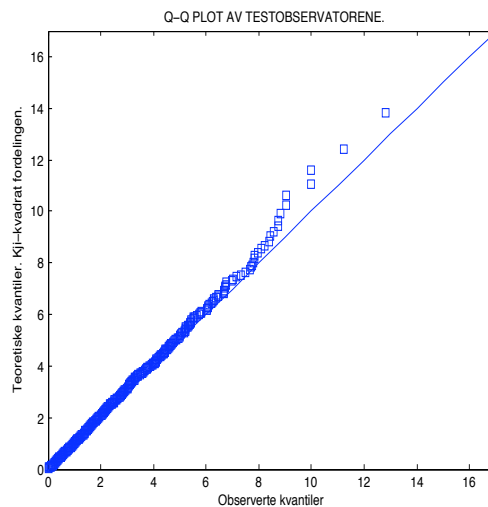


Figur 4.1: Plot av 1000 usensurerte levetider (øverst), 1000 simulerte sensureringstider (midten) og 1000 sensurerte simulerte levetider (nederst). Andel sensurering er 95%.



Figur 4.2: Histogram over estimerte regresjonskoeffisienter (øverst) og testobservatorer (nederst) etter 1000 kjøring i eksempel 8.





Figur 4.3: QQ-plot til testobservatorene i eksempel 8.

#### 4.2.2 Om tidsbruk i simuleringene

Hver kjøring i eksempel 8 brukte på min mac i gjennomsnitt omtrent 7 sekunder. De 1000 kjøringene tok dermed litt under to timer. Antall grupper som er brukt i grupperingen for testen og andel sensur har stor betydning på tidbruken i simuleringene. Når samme situasjon simuleres, men med 80% sensurering, 3 kontroller trekkes ved hvert hendelsestidspunkt og gruppering for testen i 8 grupper bruker min datamaskin i gjennomsnitt omlag 143 sekunder på hver kjøring. Dermed tar 1000 kjøringene nesten 40 timer.

### 4.3 Utvidelser av simuleringsalgoritmen

Noe av formålet med denne oppgaven er å studere teststyrken til goodness-of-fit testen (3.27) i ulike situasjoner. Når vi selv simulerer levetidene, har vi frihet til å gi dem de egenskapene vi ønsker og vi vet hva som er den riktige modellen i enhver situasjon. Ved å tilpasse feil modell kan vi dermed undersøke teststyrken i ulike situasjoner. Ved å studere plot av martingal residualprosesser når feil modell er tilpasset, kan vi også undersøke om det er mønstre i disse plottene som gjør det mulig å si hva som er feil med modelltilpassningen. I denne oppgaven vil vi undersøke testens evne til å avsløre at en modell uten tidsavhengig effekt av kovariatene er tilpasset et datasett der en slik effekt faktisk er til stede. Vi vil også undersøke teststyrken i en situasjon der en modell med lineær koding av kovariatene tilpasses data som er simulert med en log-transformert kovariat.

I dette avsnittet vil vi redgjøre for hvordan levetider med tidsavhengig effekt av kovariat og levetider med en log-transformert kovariat er simulert.

### 4.3.1 Tidsavhengig effekt av kovariatene

I medisinske studier vil man ofte observere at effekten av en eller flere forklaringsvariable varierer med tiden. Effekten røyking har på kreftrisikoen avtar for eksempel med tiden når en person slutter å røyke. Siden det ikke er usannsynlig at datamaterialet i virkelige studier inneholder tidsavhengige effekter, er det av interesse å undersøke hvordan goodness-of-fit testen (3.27) presterer i slike situasjoner. De to modellene som brukes i denne oppgaven, forutsetter begge at tidsavhengige effekter av kovariatene ikke er til stedet. Effekten av en variabel vil da estimeres til et konstant nivå som ligger mellom det høyeste og laveste nivå effekten av kovariaten virkelig tar.

I dette avsnittet følger en redgjørelse for hvordan simuleringene av levetider med tidsavhengig effekt av en kovariat er gjort i praksis. I avsnitt 4.2 så vi hvordan vi kunne simulere pseudotilfeldige levetider  $T$ , med konstant effekt av kovariatene, ved å trekke  $T$  som (4.1). Nå tar vi for oss en liknende situasjon som den i eksempel 8, men nå lar vi effekten av den uniforme kovariaten  $X_1$  være tidsavhengig. Den tidsavhengige regresjonskoeffisienten kaller vi  $\beta_1(t)$ . Vi finner da kumulativ hasard til levetidene som

$$h(t|X) = h_0(t)r((\beta_1(t), \beta_2)^T, X).$$

For Cox-modellen blir hasard og kummulativ hasard gitt ved

$$\begin{aligned} h(t|X) &= h_0(t)e^{\beta_1(t)X_1 + \beta_2 X_2} \\ H(t|X) &= \int_0^t h_0(u)e^{\beta_1(u)X_1 + \beta_2 X_2} du. \end{aligned} \quad (4.7)$$

Ved å sette at  $\beta_1(t) = \beta_1(1 + f(t))$  og bestemme en form på  $f(t)$  slik at  $f(0) = 0$ , oppnår vi at  $\beta_1(0) = \beta_1$ . Da får  $\beta_1$  en enkel tolkning, nemlig som effekten av  $X_1$  ved tid 0. Ved å sette inn for  $\beta_1(u)$  i (4.7) får vi at denne kan skrives som

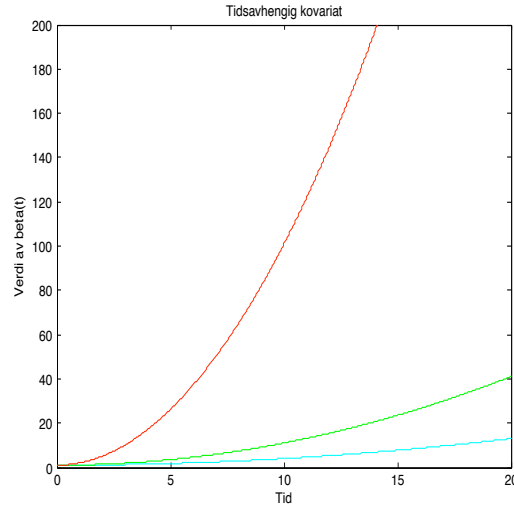
$$\begin{aligned} H(t|X) &= \int_0^t h_0(u)e^{\beta_1(1+f(u))X_1 + \beta_2 X_2} du \\ &= e^{\beta_1 X_1 + \beta_2 X_2} \int_0^t h_0(u)e^{\beta_1 f(u)X_1} du. \end{aligned} \quad (4.8)$$

Ved å velge en pen form på  $f(u)$  kan det finnes et analytisk uttrykk for  $H(t|X)$ . Vi velger  $f(t) = cH_0(t)$ , der  $H_0(t) = \int_0^t h_0(u)du$  og  $c$  er en tuningparameter. Integralet (4.8) blir dermed

$$\int_0^t h_0(u)e^{\beta_1 f(u)X_1} du = \left[ \frac{1}{c\beta_1 X_1} e^{c\beta_1 X_1 H_0(u)} \right]_0^t. \quad (4.9)$$

Vi har gitt baseline hasard  $h_0(t)$  som  $akt^{k-1}$ , og da følger det at kumulativ baseline hasard blir gitt ved  $H_0(t) = at^k$ . Da følger det at  $\beta_1(t) = (1 + cat^k)$ . Det bestemte integralet (4.9) blir dermed

$$\left[ \frac{1}{c\beta_1 X_1} e^{c\beta_1 X_1 H_0(u)} \right]_0^t = \frac{1}{c\beta_1 X_1} \left[ e^{c\beta_1 X_1 at^k} - 1 \right] \quad (4.10)$$



Figur 4.4: Plot av  $\beta_1(t)$  mot tid når tuningparameteren  $c$  tar verdien 3 (cyan), 10 (grønn) og 100 (rød) .

For  $c \neq 0$  kan vi skrive (4.8) som

$$H(t|X) = \frac{e^{\beta_1 X_1 + \beta_2 X_2}}{c\beta_1 X_1} \left[ e^{c\beta_1 X_1 a t^k} - 1 \right]. \quad (4.11)$$

Ved å invertere (4.11) finner vi at vi kan trekke levetider  $T$ , med en tidsavhengig og en tidsuavhengig kovariat, som

$$T = \sqrt[k]{\frac{\log \left[ \frac{V c \beta_1 X_1}{e^{\beta_1 X_1 + \beta_2 X_2}} + 1 \right]}{c\beta_1 X_1 a}}, \quad (4.12)$$

der  $V \sim \exp(1)$ .

Figur 4.4 viser hvordan effekten av  $X_1$  endres over tid, avhengig av verdien til tuningparameteren  $c$ . For  $c = 3$  øker den tidsavhengige regresjonskoeffisienten  $\beta_1(t)$  fra 1 ved tid 0 til omtrent 15 ved tid 20. For  $c = 10$  ser vi at  $\beta_1(t)$  tar verdien 40 ved tid 20. Valg av  $c$  lik 100 fører dermed til at  $\beta_1(t)$  vokser til omlag 400 ved samme tid.

Vi har i dette avsnittet gjort rede for hvordan vi kan simulere levetider der effekten av en kovariat er tidsavhengig. Videre har vi vist hvordan styrken på den tidsavhengige effekten kan justeres ved å endre på tuningparameteren  $c$ . I avsnittene 5.5.3 og 5.5.4 vil vi bruke teorien utviklet i dette avsnittet til å simulere levetider der effekten av en av kovariatene er tidsavhengig, for så å studere styrken til testen (3.27) i denne situasjonen.

### 4.3.2 Feil koding av kovariat

I studier kreftpassienter har det vist seg at tykkelsen på svulsten har stor innvirkning på overlevelsen til pasientene. Tykkere svulst minker sannsynligheten for å overleve. Det

viser seg imidlertid at en liten økning i svulst-tykkelse øker dødssannsynligheten mer for en pasient som hadde liten svulst til å begynne med, enn for en person som allerede hadde tykk svulst. Effekten av svulst-tykkelsen er altså ikke lineær. Effekten avtar etterhvert som tykkelsen øker. Det samme kan observeres i studier av røyking og lungekreft. Risikoen for å få lungekreft er svært mye høyere for røykere enn ikke-røykere, men effekten avtar etterhvert som konsumet av sigaretter øker. Risikoen for å utvikle lungekreft øker mer hvis en person som før røykte 10 sigaretter om dagen øker konsumet til 15, enn for en person som allerede røykte 40 sigaretter om dagen og nå øker til 45. Selv om han som nå røyker 45 sigaretter om dagen selvsagt har høyere risiko for å få lungekreft, var altså økningen i risiko større for han som nå røyker 15 sigaretter om dagen. Liknende situasjoner støter man ofte på i epidemiologiske studier.

I slike situasjoner hvor effekten av en kovariat ikke er lineær kan transformasjoner av kovariatene føre til at modellen passer bedre. I eksempelet med svulst-tykkelse ville kanskje en log-transformasjon av tykkelsen gi en bedre forståelse av virkeligheten. Andre transformasjoner kan også tenkes å være nyttige her, for eksempel kunne man tatt kvadratroten av tykkelsen.

Vi vil her kort gjøre rede for hvordan vi kan simulere levetider der effekten av en kovariat ikke er lineær, men log-transformert. Vi vil også vise hvordan vi kan justere formen på feilkodingen.

Vi husker at  $X_1$  har vært uniformt fordelt over  $[0, 1]$ . Vi ønsker nå å simulere levetider der en av kovariatene ikke er lineært kodet. Samtidig ønsker vi at utfallsrommet til kovariatene skal være det samme som før. I simuleringene erstatter vi nå denne med  $X'_1$ , som er gitt som

$$X'_1 = \frac{\log(1 + bX_1)}{\log(1 + b)}, \quad (4.13)$$

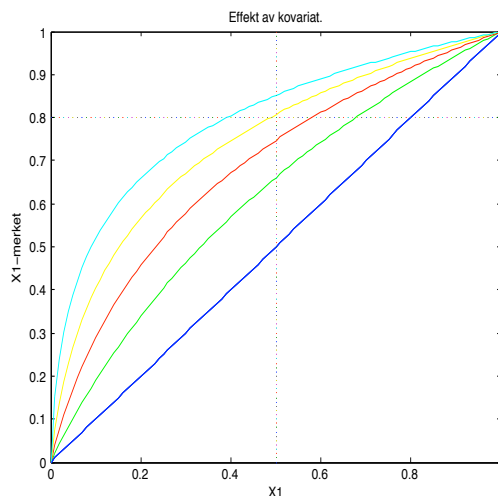
der  $X_1 \sim U(0, 1)$  og  $b$  velges slik vi får ønsket form på feilkodingen. Merk at  $X'_1 \in [0, 1]$ .

Figur 4.5 viser  $X'_1$  for ulike verdier av  $b$ . Vi ser at for høyere verdier av  $b$  blir forskjellen på  $X'_1$  og  $X_1$  større og større. For eksempel ser vi at for  $b = 30$  tar  $X'_1$  verdien 0.8 når  $X_1$  tar verdien 0.5. For  $b$  lik 3 tar  $X'_1$  omtrent verdien 0.65 for samme  $X_1$ .

I avsnitt 5.5.6 vil vi bruke metoden introdusert i dette avsnittet til å simulere levetider der effekten av en kovariat er log-transformert. Dermed vil vi kunne gjøre rede for teststyrken i situasjonen når en modell med lineær koding av kovariatene tilpasses et datasett der en kovariat er log-transformert.

## 4.4 Problemer i simuleringene

Hittil i dette kapitlet har vi redgjort for hvordan vi kan simulere levetider. Vi har også vist hvordan vi kan gi levetidene noen ønskete egenskaper, for eksempel at effekten av en



Figur 4.5: Feilkoding av kovariat. Figuren viser  $X'_1$  (4.13) for ulike verdier av  $b$ . Den rette blå streken viser  $X'_1$  som ikke-transformert. De andre strekene viser  $X'_1$  for  $b = 3$  (grønn),  $b = 10$  (rød),  $b = 30$  (gul) og  $b = 100$  (cyan).

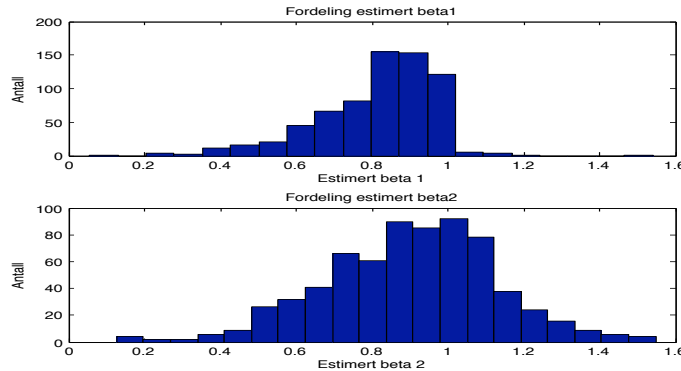
kovariat er tidsavhengig. Dette avsnittet tar for seg hvordan enkelte praktiske problemer i simuleringene er løst, og hvordan programmet er tunet for å unngå noen problemer som har oppstått underveis i simuleringene. Resultatene av simuleringene og analyser av disse kommer i neste kapittel.

#### 4.4.1 Negativ risiko

Den individuelle risikofunksjonen utgjør sammen med baseline hasard, den individuelle hasardraten (2.4) i semiparametriske regresjonsmodeller. Dermed blir tolkningen ved negativ estimert risikofunksjon at et individ har mindre estimert sannsynlighet enn 0 for å dø i neste korte tidsintervall, gitt at individet lever ved tidsintervallets begynnelse. Siden det er umulig å ha mindre sannsynlighet for å dø enn 0 bør mulige verdier av regresjonskoeffisientene begrenses slik at negativ estimert risiko ikke kan inntreffe.

I eksempel 8 ble Cox proporsjonal-hazard modell benyttet. Siden den estimerte risikofunksjonen i denne modellen er på formen  $r(\hat{\beta}, X) = e^{\hat{\beta}^T X}$  blir risikofunksjonen alltid positiv for alle individer, uansett verdier av  $X$  og  $\hat{\beta}$ . Når vi tilpasser eksess relativ risiko-modellen hvor risikofunksjonen er gitt ved  $r(\hat{\beta}, X) = (1 + \hat{\beta}_1 X_1)(1 + \hat{\beta}_2 X_2)$ , har vi ikke lenger noen garanti for positiv estimert risikofunksjon.

I denne oppgaven vil vi alltid tilpasse modeller med to kovariater. Når vi tilpasser eksess relativ risiko modellen, vil vi derfor kreve at  $(1 + \hat{\beta}_k X_k) > 0$ , for  $k = 1, 2$ . Hvilke verdier av  $\hat{\beta}_k$  som resulterer i negativ risikofunksjon avhenger av utfallsrommet til  $X_k$ . I simuleringene som er gjort i denne oppgaven er utfallsrommet til kovariatene hele tiden som i eksempel 8. Når eksess relativ risiko modellen benyttes må vi dermed kreve at verdiene av de estimerte



Figur 4.6: Histogram over  $\hat{\beta}_1$  (over) og  $\hat{\beta}_2$  (under) fra 1000 kjøringene. Eksess relativ risiko modellen er både brukt i simuleringene og tilpasset.

regresjonskoeffisientene tilfredstiller  $\hat{\beta}_1 > -1$  og  $\hat{\beta}_2 > -\sqrt{3}$ .

De nedre grensene for lovlig verdier av de estimerte regresjonskoeffisientene er så langt unna verdiene til regresjonskoeffisientene som er brukt i simuleringene at vi i praksis svært sjelden vil støte på dette problemet. I eksempel 9 vil vi likevel gjøre rede for hvordan dette kravet i andre situasjoner kan skape problemer for simuleringene våre.

**Eksempel 9 (Negativ risiko)** Dersom kovariatene kan ta både positive og negative verdier, blir begrensningene på de estimerte regresjonskoeffisientene større og lager vanskeligheter for analysene våre. I dette eksempelet ser vi på en situasjon liknende eksempel 8, med en kontinuerlig og en diktom variabel. Vi lar nå den kontinuerlige variabelen  $X_1 \sim U(-1, 1)$ , mens den diktome variabelen  $X_2$  tar verdiene  $\pm \frac{1}{\sqrt{3}}$ , med lik sannsynlighet. Da er kravet til lik varians oppfylt. Kravet til  $\hat{\beta}_1$  blir nå at denne må ligge i området  $[-1, 1]$ , mens  $\hat{\beta}_2$  må ligge i området  $[-\sqrt{3}, \sqrt{3}]$ . Effekten av en kovariat er alltid større eller lik 0 og dermed nærmer de estimerte regresjonskoeffisientene seg fort yttergrensen for gyldige verdier.

Begrensningene på de estimerte regresjonskoeffisientene fører med seg problemer når  $\hat{\beta}$  faller på utsiden av sitt lovlig område. Dersom vi ønsker at effekten av regresjonskoeffisientene skal ligge i nærheten av yttergrensene vil problemet med negativ risiko inntreffe ofte. Dersom begge regresjonskoeffisientene settes til 0.8 vil mange  $\hat{\beta}_1$  bli estimert til en ulovlig verdi og kjøringen vil stoppes. Figur 4.6 illustrerer hvordan fordelingen til  $\hat{\beta}_1$  blir skjev på grunn av begrensningene i lovlig verdier av de estimerte regresjonskoeffisientene.

I avsnittene 5.5.1 og 5.5.2 vil vi la verdien til  $\beta$  i simuleringene variere og se hvordan resultatene våre endres når regresjonskoeffisientene blir større. Denne analysen ville det være vanskeligere å gjennomføre i modellen beskrevet i dette eksemplet, da området med mulige verdier av  $\beta$  er såpass smalt. Problemet kunne løses ved å endre utfallsrommet til variablene, men det er upraktisk og måtte gjøre slike endringer hver gang regresjonskoeffisientene endres. Det er også lettere å sammenlikne resultatene av like kjøringene.

Ved å definere variablene som i eksempel 8 unngår vi begge disse problemene. Da skjer det

så sjelden at  $\hat{\beta}$  faller utenfor sitt gyldige område at det ikke er noe problem. Opprinnelig prøvde vi andre fordelinger av kovariatene. Da fikk vi slike problemer som beskrevet her.

#### 4.4.2 Divergens i scoring-algoritmen

Mens likelihooden til Cox-modellen er pen og entoppet er situasjonen en annen for eksess relativ risiko modellen. Her vet vi ingenting om hvordan likelihooden er langt unna de sanne verdiene. Når vi skal estimere regresjonskoeffisientene til Cox modellen med Fischer-scoring algoritmen (3.10) er vi derfor trygge på at denne vil konvergere til den verdien av  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$  som maksimerer likelihooden (3.4). For eksess relativ risiko modellen er vi derimot ikke sikre på at dette vil skje. Vi vet at det vil finnes et lokalt maksimum i området i nærheten av de sanne regresjonskoeffisientene, men av ulike årsaker kan algoritmen i enkelte tilfeller unngå å konvergere til dette punktet.

Vi vet videre at de partielt deriverte til log-likelihoodfunksjonen, score funksjonene (3.9), skal være tilnærmet lik null i et toppunkt. Ved å legge dette til Fisher-scoring algoritmens konvergenskriterium  $|\hat{\beta}_{OLD} - \hat{\beta}_{NEW}| < \epsilon$ , kan man forsikre seg om at punktet algoritmen har konverget til faktisk er et toppunkt.

Hvis scoring algoritmen kommer på utsiden av området som inneholder maksimumspunktet vi er ute etter, kan scoring algoritmen divergere, konvergere til et annet lokalt maksimum, eller, avhengig av kriteriene som velges for å avslutte algoritmen, avslutte med en verdi som ikke er et lokalt maksimum. Avgjørende for hvorvidt algoritmen hopper ut av det ønskede området er skrittlengden  $\delta$ , det vil si hvor langt algoritmen hopper ved hver iterasjon. Dersom denne er for stor til å begynne med, kan algoritmen hoppe ut av området avgrenset av de verdier av  $\hat{\beta}$  som gir meningsfulle verdier, det vil si at de tilfredstiller kravet om positiv estimert risikofunksjon. Løsningen blir en modifisering av Fischer-Scoring algoritmen (3.10). Den nye algoritmen tar formen

$$\beta^{(s+1)} = \beta^{(s)} + \delta^{(s)} I^{-1}(\beta^{(s)}) s(\beta^{(s)}), \quad (4.14)$$

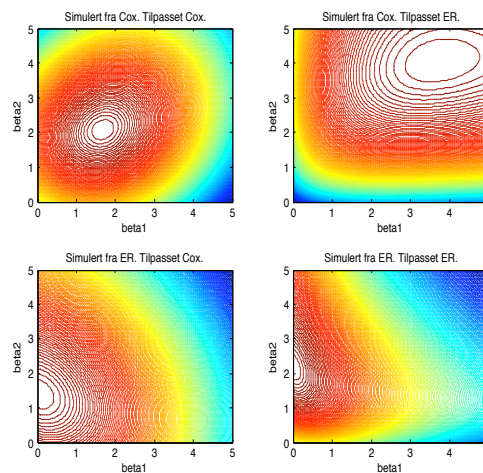
der  $\delta^{(s)}$  er skrittlengden ved  $s$ -te iterasjon. Skrittlengden ble etter noe prøving satt til 0.3 en verdi som viste seg å virke bra med de parameterverdiene som er valgt.

Dersom algoritmen likevel konvergete til en ulovlig verdi, konvergete til en verdi der score funksjonene ikke var tilnærmet lik 0, eller konvergens ikke var oppnådd etter 200 iterasjoner, ble scoring algoritmen startet på nytt, med samme startverdier som sist, men med halvert skrittlengde. Dette ble gjentatt inntil konvergens var oppnådd eller skrittlengden ble mindre enn  $s_{min}$ , som ble satt til 0.1. Dersom konvergens ikke var oppnådd da, ble datsettet forkastet. Uten estimerte regresjonskoeffisienter er det ikke mulig å utføre testen og kjøringen ble dermed avbrutt.

**Eksempel 10 (Divergens i scoring algoritmen)** For å illustrere problemene med å få scoring algoritmen til å konvergere til et gyldig lokalt toppunkt, er det tatt med fire konturplott i figur 4.7. Disse viser hvordan tilpasningen til Cox og eksess relativ risiko - modellen blir på to tilfeldige datasett, der det ene er simulert med Cox-modellen, mens det

Tabell 4.1: De estimerte regresjonskoeffisientene for begge risikofunksjonene tilpasset datasett simulert med begge risikofunksjonene.

Modell simulert	Modell tilpasset	$\hat{\beta}_1$	$\hat{\beta}_1$
Cox	Cox	1.6701	2.0769
Cox	ER	3.6966	4.0358
ER	Cox	0.0174	1.4008
ER	ER	0.0191	2.1555



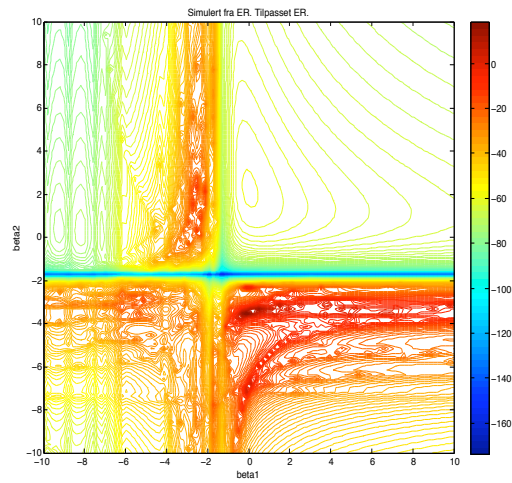
Figur 4.7: Konturplot av logaritmisk likelihood til tilpasningen av riktig og feil modell til datasett simulert fra Cox-modellen og eksess relativ risiko modellen.

andre er simulert med eksess relativ risiko modellen. De estimerte regresjonskoeffisientene fra den modifiserte scoring-algoritmen (4.14) er oppgitt i tabell 4.1. Vi ser at algoritmen har funnet toppunkter i hver av de fire situasjonene i figur 4.7 og at den logaritmiske likelihooden ser ut til å være pen omkring de estimerte verdiene. Denne vil imidlertid variere fra datasett til datasett.

Figur 4.8 viser den logaritmiske likelihooden i et større område når eksess reativ risiko modellen både er simulert fra og tilpasset. Dette tilsvarer plottet nederst til høyre i figur 4.7. Det er lett å kjenne igjen toppunktet funnet av scoring algoritmen (0.0191, 2.1555), men dette plottet viser også at dette bare er et lokalt toppunkt siden det finnes områder i nærheten der den logaritmiske likelihooden er høyere. I dette plottet svarer dette til områder der risikofunksjonen ville få negativ verdi. Hvis skrittlengden til scoring-algoritmen var for stor kunne denne hoppet ut i disse områdene og vi hadde fått et ugyldig estimat for  $\beta$ .

I dette kapittelet har vi vist hvordan man kan simulere levetider med en ønsket form på kumulativ hasard og ønsket andel sensurering. Vi har også vist hvordan vi kan simulere





Figur 4.8: Konturplot av logaritmisk likelihood til tilpasningen av eksess relativ risikomodellen når data er estimert fra samme modell.

levetider med ulike egenskaper, slik som tidsavhengig effekt av en kovariat eller feil koding av en kovariat. Til slutt har vi tatt for oss noen problemer som har oppstått i simuleringene og hvordan disse er løst. Hensikten med dette har vært å utvikle simuleringsverktøy som gjør oss i stand til å studere egenskapene til goodness-of-fit testen (3.27) i ulike situasjoner. I neste kapittel vil vi gjøre rede for hvilke simuleringer som er gjort, og analysere resultatene av disse simuleringene.



## Kapittel 5

# Simuleringsresultater

### 5.1 Innledning

I kapittel 4 er det redgjort for hvordan simuleringene er gjort og hvorfor vi endte med å simulere data på den formen vi gjorde. I dette kapitlet vil vi oppsummere og analysere resultatene fra simuleringene som er gjort. Det vil bli redgjort for hvilke faktorer som vil varieres i simuleringene og hvordan goodness-of-fit testen (3.27) vil brukes. Vi vil analysere signifikansnivået til testen og teststyrken i ulike situasjoner.

I eksempel 5 så vi hvordan vi kunne tilpasse en regresjonsmodell til et virkelig datasett. I eksempel 6 og 7 gikk vi et skritt videre og så hvordan vi kunne undersøke modelltilpasningen ved hjelp av grupperte martingal residualprosesser og en kji-kvadrat goodness-of-fit test (3.27). I eksempel 6 konkluderte vi med at hverken Cox-modellen eller eksess relativ risiko modellen passet godt til data, men det var vanskelig å svare på hva den dårlige tilpasningen skyldtes.

Ved å analysere simulerte levetidsdata kan vi få innsikt i hvordan analyseverktøyene våre fungerer. Her kan vi gi data de egenskaper vi ønsker og studere hvordan modellene er egnet til å analysere data med ulike karakteristikk. For eksempel kan vi studere hva slags resultater vi får i Cox-modellen dersom vi har tidsavhengig effekt av kovariatene. Hvordan andelen sensurerte levetider påvirker resultatene kan også være av interesse. Vi vil hovedsakelig bruke simuleringene til å studere egenskapene til goodness-of-fit testen (3.27). Hvordan storutvalgsegenskapene til denne testen holder i forskjellige situasjoner vil være av spesiell interesse.

### 5.2 Karakteristikker ved data

I dette avsnittet vil vi redgjøre for hvilke situasjoner som er undersøkt og hvorfor. Fremgangsmåten beskrevet i eksempel 8 er brukt med tanke på hvordan levetider og sensureringstider er trukket. I dette eksemplet ble det til hvert kasus samlet en kontroll ved

enkel nøstet kasus kontroll. Andelen sensur var 95%. Hvorvidt testens nivå og styrke blir påvirket av disse karakteristikkene ved data, er noen av tingene vi ønsker å undersøke i denne oppgaven. Derfor er simuleringene gjentatt, mens egenskapene til data er endret. Hva som er endret og på hvilke nivåer presenteres i dette avsnittet.

For å undersøke hvordan testen (3.27) og resten av maskineriet fungerer i forskjellige situasjoner, ble det valgt fire forskjellige faktorer som er variert. Ved siden av disse fire faktorene ønsker vi å undersøke om grupperingen som er brukt i testen har betydning for signifikansnivå eller teststyrke i ulike situasjoner. Mer om dette i avsnitt 5.3. De fire faktorene vi vil variere ellers er:

- Risikofunksjon: Cox semiparametriske regresjonsmodell eller Eksess relativ risiko modellen.
- Andel sensurering: 80% , 90% eller 95%.
- Samplingdesign: Enkel nøstet kasus kontroll eller Counter-Matching.
- Antall kontroller samlet ved hvert hendelsestidpunkt: 1 eller 3.

Flere faktorer kunne være interessante å undersøke, slik som flere kovariater, flere risiko-funksjoner, forskjellig gruppering ved Counter-Matching mm. Vi måtte likevel begrense oss på grunn av begrenset tid og begrenset antall datamaskiner til rådighet. Faktorene ovenfor ble ansett som mest interessante. Spesielt vil det være nyttig å finne ut om antall kontroller eller samplingdesign har stor betydning da denne kunnskapen kan brukes direkte ved design av studier i fremtiden. Det samme gjelder kunnskap om teststyrken til de ulike grupperingene for testen.

For det første er det av interesse om valg av risikofunksjon kan ha betydning. Vi har valgt å begrense oss til to modeller, nemlig Cox semiparametriske regresjonsmodell, og eksess relativ risiko modellen, som begge er definert tidligere i oppgaven. Ved å ha to modeller å velge mellom får vi også muligheten til å simulere fra en modell og tilpasse en annen.

Andel sensurering har betydning for presisjonen til de estimerte regresjonskoeffisientene. Om det også har betydning for testen, er en av tingene vi er interessert i å studere. Ved å la andelen sensur variere mellom tre nivåer, 80%, 90% og 95% henholdsvis, kan vi studere hvilke konsekvenser dette får for analysen. Disse tre nivåene er valgt fordi de er realistiske i forhold til virkelige kasus-kontroll studier.

De to samplingdesignene presentert i kapittel 3, nemlig enkel nøstet kasus-kontroll samt Counter-Matching er begge brukt i simuleringene. Vi har tidligere sett hvordan Counter-Matching gir en mer informativ samlet kontrollmengde som kan gi oss mer presise estimater på regresjonskoeffisientene. Hvordan dette vil innvirke på testen kan vi undersøke ved å prøve begge de to metodene. Ved Counter-Matching er individene delt i to grupper etter en kovariat, nemlig hvorvidt  $X_1$  er større eller mindre enn 0.5. Dette gjelder for alle simuleringene.

Antall kontroller samlet ved hvert tidspunkt er variert mellom 1 og 3. Når kun en kontroll samples blir den samlede risikomengden naturlig nok mindre informativ. Ved Counter-

Tabell 5.1: Testobservatorer fra hver av de 24 kombinasjonene i eksempel 11.

Modell	Eksess Relativ Risiko				Cox			
Samplingdesign	CM		KK		CM		KK	
Antall Kontroller	1	3	1	3	1	3	1	3
80% sensur	2.55	0.92	1.73	2.69	2.95	3.35	0.74	2.09
90% sensur	1.55	0.66	0.17	0.01	2.26	3.47	4.12	2.50
95% sensur	0.12	1.09	0.81	0.04	2.03	2.93	2.08	1.38

Matching er det i tilfellet med en kontroll per kasus samplet blandt individer som er under risiko, men som ikke er i samme stratum som kasus. Når det samples tre kontroller per kasus ved Counter-Matching trekkes en kontroll fra samme stratum som kasus, mens to kontroller trekkes fra det andre stratum.

Til sammen har vi 4 faktorer som til sammen kan kombineres på 24 forskjellige måter. I neste avsnitt vil vi presentere ulike grupperinger for testen (3.27) For hver av disse grupperingene er hver av de 24 situasjonene beskrevet i dette avsnittet brukt.

**Eksempel 11 (De 24 kjøringene)** I dette eksempelet vil vi illustrere hvordan simuleringresultatene påvirkes av de fire faktorene beskrevet over. Ved å gjenta simuleringene gjort i eksempel 8 for hver av de 24 kombinasjonene av de fire faktorene, kan vi få et første inntrykk av hvordan disse faktorene påvirker simuleringresultatene.

For hver av de 24 kombinasjonene av risikofunksjon, sensureringsandel, antall kontroller og samplingdesign som er beskrevet over, ble det trukket 1000 levetider etter samme oppskrift som i eksempel 8. Levetidene ble dermed simulert fra Cox-modellen halvparten av gangene og eksess relativ risiko modellen den andre halvparten. Individenes kovariatverdier var de samme i alle tilfellene. Sensureringstider ble trukket slik at ønsket andel ble sensurert. Ved hvert hendelsestidspunkt  $t$  ble det trukket 1 eller 3 kontroller etter enkel nøstet kasus kontroll eller Counter-Matching. I hver situasjon ble den sanne modellen tilpasset data, og til slutt ble goodness-of-fit testen (3.27) utført. Grupperingen i testen ble gjort i to tidgrupper og to kovariatgrupper, som i eksempel 7.

Testobservatorene er gitt i tabell 5.1. Det kan se ut til at testobservatorene blir noe høyere når kontrollene samples ved Counter-Matching enn ved enkel nøstet kasus kontroll. Ellers er det vanskelig å se mønstre i tabellen.

Siden samme gruppering er brukt i testen i hver av de 24 situasjonene, og den sanne modellen er tilpasset, skal testobservatorene være  $\chi^2$ -kvadrat fordelt med to frihetsgrader. Selv om det er en del variasjon mellom testobservatorene er det ingen av dem som gir forkastning på 5% eller 10% nivå.

## 5.3 Grupperinger i testen

I forrige avsnitt tok vi for oss hvilke ulike faktorer ved de simulerte levetidsdatasettene vi vil variere for å underøke effekten av i forhold til styrke og nivå på testen. I dette avsnittet tar vi for oss hvordan testen kan gjøres på ulike måter, ved å gjøre grupperingen av martingal residual prosessene etter forskjellige kriterier, og hvordan dette kan påvirke testens nivå og styrke.

Vi husker fra avsnitt 3.7 at kji-kvadrat goodness-of-fit testen (3.27) baserer seg på å sammenlikne antall observerte og forventede hendelser i bestemte grupper. Hvilke variable det skal grupperes etter eller hvor mange grupper det skal deles inn i er opp til den som skal gjøre analysen. Et av hovedmålene med denne oppgaven har vært å belyse hvorvidt det har betydning for signifikansnivå eller styrke til testen hvordan denne grupperingen gjøres. Vi har variert antall grupper det skal deles inn i, men også hvilke variable grupperingen skal gjøres etter. En mulighet er å dele inn etter tid og kovariat som i eksempel 8, mens en annen mulighet er å erstatte kovariatet med risikofunksjonen til individene. En tredje mulighet blir å dele inn kun etter risiko. Andre muligheter kan også tenkes, men vi har holdt oss til disse situasjonene i denne oppgaven da disse ble ansett som mest interessante.

Kji-kvadrat goodness-of-fit tester baserer seg gjerne på å analysere forskjellen på observert og estimert antall hendelser for grupper av individer. Antall grupper det lønner seg å dele individene inn i i forbindelse med slike tester er gjenstand for diskusjon. En tommelfingerregel sier at hver gruppe bør ha minst fem forventede observasjoner (Rice, 1995). Siden vi ønsker å bruke den samme grupperingen for de tre gradene av sensurering vi vil veksle mellom i denne oppgaven, må vi ta utgangspunkt i situasjonen med mest sensurering og da har vi i gjennomsnitt 50 kasus. Denne regelen tillater da maksimalt 10 grupper, men da må disse tilpasses hver kjøring for at kravet skal være oppfylt ved at hver gruppe inneholder 5 forventede kasus.

En av variablene vi vil bruke for gruppering i testen er individenes risikofunksjon. Dette kan være fordelaktig framfor gruppering etter en kovariat da også individets andre kovariatverdier, samt vektingen av risikoen disse utgjør ved den estimerte risikofunksjonen, blir tatt i betraktning. Ved gruppering etter individenes risiko grupperes individene enten ved kvartiler eller kvantiler av risikofunksjonene til kasus eller kasus og kontroller. Selv om vi simulerer levetiden til 1000 individer og dermed kjenner kovariatverdien til alle disse er det ikke tilfellet i en kasus-kontroll studie. Derfor kan vi ikke bruke informasjon om risikofunksjonen til alle individer i kohorten ved gruppering. I en virkelig kasus-kontroll studie har man derimot informasjon om alle kasus og alle kontroller. Individer kan opptre både som kasus og kontroller, og kontroller flere ganger, men vi vil se bort fra dette og gruppere etter kvartiler eller kvantiler av alle disse risikofunksjonene. Siden kasusene gjerne har høyere risiko enn kontrollene betyr den første måten å gruppere på gjerne at gruppene med lavest risiko vil inneholde færre kasus enn gruppene med høyere risiko. Dersom det i tillegg deles inn i mange grupper kan dette føre til at kravet om fem observasjoner i hver gruppe ikke blir overholdt. Ved å gruppere etter kun kasusenes risiko er vi garantert å få like mange observasjoner i hver gruppe, og dermed kan kravet om fem observasjoner i hver gruppe enkelt overholdes ved å bestemme et lite nok antall grupper.

En simuleringsstudie av Hosmer & Lemeshow (1989) viste at ved logistisk regresjonsanalyse av en binomisk situasjon, var det fordelaktig å gruppere etter persentiler framfor faste grupper. Dette gjaldt spesielt når mange av observasjonene havnet i enkelte grupper. Angående valg av antall grupper, gjør studien til Hosmer og Lemeshow også noen bemerkninger. Når færre enn 6 grupper ble brukt konkluderte goodness-of-fit testen nesten alltid med at det var riktig modell som var tilpasset. I denne oppgaven vil testen utføres både med færre (4) og flere (8) enn 6 grupper. Vi får derfor muligheten til å undersøke om denne konklusjonen kan overføres til vår situasjon.

En ulempe med å dele inn etter risiko er at vi bryter kravet om at man ikke skal benytte seg av informasjon man får fremover i tid, når man grupperer. Her vil grensene for de ulike gruppene nettopp være kvartiler eller percentiler av risikofunksjonen til alle kasus eller alle kasus og samplete kontroller. Når vi skal plassere kasus ved tid  $t$  i en gruppe etter den individuelle risikofunksjonen, bruker vi dermed informasjon om de individene som dør etter tid  $t$ . Likevel håper vi at dette bruddet på forutsetningene for metoden ikke medfører større problemer for oss enn at null-fordelingen til testen blir tilnærmet den samme. Dette er ett av momentene vi vil undersøke med simuleringene våre.

Til sammen har vi kombinert de ovennevnte faktorene som varieres ved gruppering i testen på syv ulike måter. Man kunne selvsagt prøvd flere kombinasjoner, men disse ble ansett som mest interessante. På grunn av begrensede ressurser ville det også ta lang tid å prøve flere kombinasjoner. De syv designene som er valgt for gruppering i testen er:

1. To kovariatgrupper, to tidgrupper (2Tid2Kov).
2. Fire grupper etter risikofunksjon for kasus og kontroller (4RiskR).
3. Åtte grupper etter risikofunksjon for kasus og kontroller (8RiskR).
4. Fire grupper etter risikofunksjon for kasusene (4RiskKasus).
5. Åtte grupper etter risikofunksjon for kasusene (8RiskKasus).
6. To tidgrupper og to grupper etter risikofunksjon for kasus og kontroller (2Tid2RiskR).
7. To tidgrupper og fire grupper etter risikofunksjon for kasus og kontroller (2Tid4RiskR).

Forkortelsene i parentes vil for enkelhets skyld brukes i tabeller og noen ganger i diskusjon senere.

**Eksempel 12 (Grupperingseksempel)** I eksempel 7 så vi på levetidene til urangruvearbeiderne fra Colorado. Ulike modeller ble tilpasset til datasettet og til slutt sjekket vi modelltilpasningen med goodness-of-fit testen (3.27). Her ble individene delt inn i seks grupper, to tidgrupper, og tre grupper etter radoneksponering. Når eksess relativ risiko-modellen ble tilpasset og tre kontroller ble samlet per kasus ved enkel nøstet kasus kontroll, tok testobservatoren verdien 10.5. Sammeliknet med  $\chi^2$ -kvadrat fordelingen med 4 frihetsgrader ga dette forkastning på 5%-nivået.

Tabell 5.2: Testobservatorer når eksess risk modellen er tilpasset datasett fra eksempel 7. Tre kontroller er trukket per kasus ved enkel nøstet kasus kontroll, og de syv ulike designene for gruppering i testen presenter i avsnitt 5.3 er brukt. Forklaring til forkortelsene i tabellen er gitt i teksten.

Gruppering	$\chi^2$	df	p-verdi
1) 2Tid2Kov	11.7	2	0.003
2) 4RiskR	6.7	3	0.082
3) 8RiskR	15.5	7	0.030
4) 4RiskKasus	5.0	3	0.172
5) 8RiskKasus	13.7	7	0.057
6) 2Tid2RiskR	1.7	2	0.427
7) 2Tid4RiskR	13.4	6	0.037

For å få bedre følelse med hvordan de ulike grupperingene for testen fungerer, er testen utført med de syv ulike grupperingene på det samme datasettet. Der det skulle grupperes etter tid ble grensen som i eksempel 7 satt til over eller under 60 år. Ved inndeling etter kovariat ble radoneksponering valgt. Goodness-of-fit observatoren fra de syv kjøringene er gitt i tabell 5.2.

Vi ser at testobservatorene varierer mye avhengig av hvordan grupperingen i testen er gjort. Den første situasjonen, der individene er delt inn i to tidgrupper og to kovariatgrupper, gir forkastning på et hvert rimelig nivå. Testen gir forkastning på 5%-nivå for tre av grupperingene. I de fire resterende tilfellene gir testen ikke forkastning. Hvorvidt disse resultatene lar seg generalisere vil vi undersøke i de neste avsnittene.

## 5.4 Når modellen stemmer

I dette avsnittet vil vi undersøke om signifikansnivået til goodness-of-fit testen (3.27) holder. Samtidig vil vi introdusere de ulike verktøyene vi har til å analysere simuleringresultatene. I eksempel 7 visste vi ikke om riktig modell var valgt. Testen ga svært ulike resultater for Cox og eksess relativ risiko modellen. Selv om testen ga forkastning for begge modellene på 1% nivå, ble ikke eksess relativ risiko modellen forkastet da kontroller ble trukket med Counter-Matching.

Når vi simulerer er vi i den situasjonen at vi vet hvilken modell som faktisk er den rette. Ved å tilpasse den sanne modellen og deretter utføre goodness-of-fit testen, kan vi undersøke om antagelsene om fordelingen til testobservatorene under nullhypotesen stemmer. Fra avsnitt 3.7 husker vi at fordelingen til testobservatorene under nullhypotesen skal være  $\chi^2$ -kvadrat fordelt med  $K(G - 1)$  frihetsgrader. Her er  $G$  antall grupper individene deles inn i etter kovariat eller risiko, mens  $K$  er antall tidgrupper.

Også i dette avsnittet har vi brukt samme proseyre som i eksempel 8. Ved å gjenta denne



Tabell 5.3: Når modellen stemmer: Antall ganger scoring-algoritmen ikke konvergente. Gruppering i fire grupper: 2 tidgrupper og 2 kovariatgrupper. Her er Eksess relativ risiko modellen og Cox-modellen brukt. Samplingdesignene er Enkel nøstet kasus kontroll (KK) og Counter-Matching (CM). Antall kontroller trukket til hvert kasus er 1 eller 3, men forventet andel sensurerte levetider varierer mellom 0.80, 0.90 og 0.95.

Modell	Eksess Relativ Risiko				Cox			
Samplingdesign	CM		KK		CM		KK	
Antall kontroller	1	3	1	3	1	3	1	3
80% sensur	0	0	0	0	0	0	0	0
90% sesnur	0	0	0	0	0	0	0	0
95% sensur	14	9	24	10	0	0	0	0

prosedyren 1000 ganger for hver kombinasjon av de fire faktorene beskrevet i avsnitt 5.2, kan vi undersøke om noen av disse påvirker signifikansnivået til testen. Ved å, for hver av disse 24 kombinasjonene gjøre testen ved hver av de syv grupperingene presentert i avsnitt 5.3, kan vi videre undersøke om signifikansnivået holder i hver av disse situasjonene. Siden vi tester alle kombinasjoner av de 24 situasjonene mot hver av de syv grupperingene, kan vi videre undersøke om det finnes interaksjonseffekter mellom disse. Vi vil først oppsummere simuleringsresultatene i noen enkle tabeller som viser andel kjøringene der scoring-algoritmen ikke konvergente, samt andel av kjøringene der testen ga forkastning. Deretter vil vi se nærmere på resultatene gjennom en regresjonsanalyse av disse.

Først vil vi se nærmere på simuleringsresultatene når det i testen er delt i fire grupper, to kovariatgrupper og to tidgrupper. Tabell 5.3 viser antall ganger Scoring-algoritmen ikke konvergente når riktig modell ble tilpasset. Likelihooden til Cox-modellen er entoppet og det er derfor ikke overaskende at scoring-algoritmen alltid konvergerer når denne modellen tilpasses. Også når eksess relativ risiko modellen tilpasses konvergerer algoritmen i de aller fleste situasjonene. Vi ser også at algoritmen konvergerer oftere når kontrollene samples ved Counter-Matching enn ved enkel nøstet kasus kontroll. Den siste konklusjonen vi kan trekke fra tabell 5.3 er at algoritmen konvergerer oftere når det samples tre en kun en kontroll per kasus. Sammen med Counter-Matching fører tre kontroller nødvendigvis til en mer informativ kontrollgruppe. Jo mere data vi har, jo penere blir likelihooden i et område rundt de sanne regresjonskoeffisientene, og dermed blir det lettere for scoring-algoritmen å finne dette punktet.

Det er vanskelig å se om nivået til testen (3.27) avhenger av noen av de fire faktorene nevnt i forrige avsnitt. Tabell 5.4 viser at nivået ligger i nærheten av 5% for de aller fleste kjøringene. Situasjonen med høyest nivå er når vi bruker Cox-modellen, 1 kontroll samples ved hvert hendelsestidspunkt ved enkel nøstet kasus kontroll og andelen sensur er 95%. Her gir testen forkastning i 7.0% av kjøringene. Når Cox-modellen brukes, 3 kontroller samples ved hvert hendelsestidspunkt ved Counter-Matching og andelen sensur er 80% eller 95%, får vi det lavest estimerte gjennomsnittlige signifikansnivået. Her gir testen forkastning i kun 4.0% av kjøringene. Alle de syv grupperingene er kjørt på samme måte som her. Simuleringsresultatene fra disse er oppsummert i tabell A.1 i vedlegget. Det er

Tabell 5.4: Andel tester i prosent som gir forkastning. Gruppering: 2Tid2Kov. Riktig modell tilpasset.

Modell	Eksess Relativ Risiko				Cox			
Samplingdesign	CM		KK		CM		KK	
Antall kontroller	1	3	1	3	1	3	1	3
80% sensur	5.3	4.9	4.8	4.5	5.6	4.0	4.7	4.5
90% sensur	4.8	4.7	5.6	5.0	5.2	6.0	6.1	4.8
95% sensur	4.4	4.5	5.4	5.4	5.6	4.0	7.0	5.9

svært vanskelig å se noen mønstre eller trekke noen konklusjoner etter kun å ha sett på tabell, men nivået ser ut til å holde godt.

#### 5.4.1 Regresjonsanalyse av signifikansnivået

Regresjonsanalyse av simuleringsresultatene er en effektiv måte å få oversikt over hvilke faktorer som påvirker signifikansnivået, og senere teststyrken til goodness-of-fit testen (3.27). I dette avsnittet vil vi ved hjelp av regresjonsanalyse av simuleringsresultatene se nærmere på om noen faktorer påvirker signifikansnivået til goodness-of-fit testen. Vi vil først gi en kort beskrivelse av hvordan logistisk regresjon kan brukes til å analysere simuleringsresultatene, og deretter se på hovedresultatene av denne analysen.

Vi definerer at forkastning er suksess. For hver kombinasjon av de 24 kjøringene og 7 grupperingene, blir antall suksesser binomisk fordelt. Til sammen har vi dermed 24 binomiske utvalg der  $Y_i \sim B(n_i, p_i)$ . Siden responsen, altså antall forkastninger, er binomisk fordelt brukes logistisk regresjon. Vi tilpasser da en regresjonsmodell på formen

$$\log \left( \frac{p_i}{1 + p_i} \right) = \alpha + \beta_r + \gamma_s + \delta_d + \epsilon_n + \theta_q,$$

der  $\alpha$  er konstantledd,  $\beta_r$  er effekten av  $r$ -te gruppering,  $\gamma_s$  er effekten av  $s$ -te risiko-funksjon,  $\delta_d$  er effekten av  $d$ -te andel sensurering,  $\epsilon_n$  er effekten  $n$ -te samplingdesign og  $\theta_p$  er effekten av  $p$ -te antall kontroller. Vi har brukt sum-parametrisering som for eksempel betyr at  $\sum_{r=1}^7 \beta_r = 0$ . Dermed kan vi estimerte det gjennomsnittlige signifikansnivået ved  $\frac{e^{\hat{\alpha}}}{1+e^{\hat{\alpha}}}$ . De resterende koefisientene forteller oss hvilken effekt de ulike faktorene har på signifikansnivået i forhold til det gjennomsnittlige nivået. Et estimat for det gjennomsnittlige signifikansnivået til simuleringene med 80% sensurering finner vi for eksempel som  $\frac{e^{\hat{\alpha}+\hat{\delta}_1}}{1+e^{\hat{\alpha}+\hat{\delta}_1}}$ , der  $\hat{\delta}_1$  er den estimerte effekten av 80% sensur.

Vi er vant til at sannsynligheten for suksess typisk vil avhenge av en eller flere av forklaringsvariablene. Her er forklaringsvariablene de faktorene vi varierer i simuleringene, presentert over. Vi undersøker altså om en eller flere av faktorene har betydning for signifikansnivået til testen. Vi lar også hvilken gruppering som er brukt i testen være en forklaringsvariabel i regresjonsanalysen. På den måten kan vi undersøke hvorvidt nivået

Tabell 5.5: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Når modellen stemmer.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
$\alpha$ ) Konstantledd	-2.886	0.011	-262.328	$< 2 \cdot 10^{-16}$
$\beta_1$ ) 2Tid2Kov	-0.037	0.027	-1.379	0.168
$\beta_2$ ) 4RiskR	-0.009	0.027	-0.328	0.743
$\beta_3$ ) 8RiskR	-0.052	0.027	-1.896	0.058
$\beta_4$ ) 4RiskKasus	0.042	0.026	1.579	0.114
$\beta_5$ ) 8RiskKasus	0.177	0.025	7.035	$1.99 \cdot 10^{-12}$
$\beta_6$ ) 2Tid2RiskR	-0.009	0.027	-0.328	0.743
$\gamma_1$ ) Cox	0.009	0.011	0.791	0.429
$\delta_1$ ) 80%c	-0.032	0.016	-2.078	0.038
$\delta_2$ ) 90%c	0.002	0.015	0.097	0.923
$\epsilon_1$ ) KasusKontroll	-0.011	0.011	-1.040	0.298
$\theta_1$ ) 1Kontroll	-0.063	0.011	-5.756	$8.63 \cdot 10^{-9}$

holder i alle de syv situasjonene, eller om det er forskjeller mellom disse. Til slutt i avsnittet vil vi også undersøke om det finnes interaksjonseffekter av 1. orden mellom noen av disse som er signifikante på 1%-nivået.

Tabell 5.5 viser estimerte regresjonskoefisienter, standardfeil, z-verdi og tilhørende p-verdi for hvert nivå av faktorene som varieres. Siden det er valgt at summen av koefisientene til hver faktor skal være 0, kan vi finne estimatet for det siste nivået til hver faktor som summen av estimatene til de andre nivåene til samme faktor, multiplisert med  $-1$ .

Det gjennomsnittlige signifikansnivået for testen estimeres av modellen uten interaksjonseffekter til  $\frac{e^{\hat{\alpha}}}{1+e^{\hat{\alpha}}} = 0.053$ .

Som vi ser av tabell 5.5 er den femte grupperingen ved testen, og antall kontroller som samples, de eneste faktorene som har effekt på signifikansnivået som er signifikante på 1% nivå. Den femte grupperingen ved testen er når individene deles inn i åtte grupper etter risikofunksjonen til kasusene. Siden den estimerte regresjonskoefisienten her er 0.177, betyr dette at nivået blir høyere med denne grupperingen. Nærmere bestemt estimeres det gjennomsnittlige signifikansnivået i denne situasjonen til  $\frac{e^{\hat{\alpha}+\hat{\beta}_5}}{1+e^{\hat{\alpha}+\hat{\beta}_5}} = 0.062$ .

For antall kontroller ser vi at når det kun samples 1 kontroll pr kasus, blir nivået lavere enn når det samples 3 kontroller pr kasus. Ved sampling av kun 1 kontroll blir nivået estimert til 0.050. Når 3 kontroller samples blir det gjennomsnittlige nivået 0.056 når interaksjonseffekter ikke er med. De resterende faktorene samplingdesign, risikofunksjon og andel sensurering har ikke betydning for det gjennomsnittlige signifikansnivået til testen.

Når modellen tilpassses med alle 1. ordens interaksjonseffekter er det bare interaksjonseffekter mellom gruppering i testen og antall kontroller som samples som blir signifikante på 1% nivå. Resultatene når vi tilpasser modellen uten ikke-signifikante interaksjoner er gitt i tabell 5.6.

Tabell 5.6: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Når modellen stemmer.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.943	0.011	-267.331	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.032	0.027	-1.184	0.237
4RiskR	-0.004	0.027	-0.150	0.881
8RiskR	-0.049	0.028	-1.778	0.076
4RiskKasus	0.044	0.026	1.655	0.098
8RiskKasus	0.150	0.026	5.842	$5.15 \cdot 10^{-9}$
2Tid2RiskR	-0.004	0.027	-0.149	0.882
Cox	0.008	0.011	0.739	0.460
80%c	-0.030	0.016	-1.938	0.053
90%c	0.002	0.015	0.103	0.918
KasusKontroll	-0.011	0.011	-0.984	0.325
1Kontroll	-0.057	0.011	-5.140	$2.75 \cdot 10^{-7}$
2Tid2Kov:1Kontroll	0.108	0.027	3.983	$6.80 \cdot 10^{-5}$
4RiskR:1Kontroll	0.029	0.027	1.075	0.282
8RiskR:1Kontroll	-0.033	0.028	-1.211	0.226
4RiskKasus:1Kontroll	0.030	0.026	1.142	0.254
8RiskKasus:1Kontroll	-0.151	0.026	-5.895	$3.75 \cdot 10^{-9}$
2Tid2RiskR:1Kontroll	0.032	0.027	1.195	0.232

I regresjonsanalysen som inkluderer de signifikante interaksjonene bekreftes resultatene fra regresjonsanalysen uten interaksjoner. Det gjennomsnittlige signifikansnivået til testen estimeres til 0.050. Fortsatt blir situasjonen der grupperingen til testen er åtte grupper etter risikofunksjonen til kasusene samt antall kontroller signifikante på 1% nivå. Her estimeres nivået til testen til 0.058 og 0.047 henholdsvis. Av 1. ordens interaksjoner er det bare interaksjoner mellom hvilken gruppering som benyttes til testen samt antall kontroller som blir signifikante. Ingen interaksjonseffekter er likevel særlig store. Samplingdesign, risikofunksjon, eller andel sensurering ser ikke ut til å ha betydning for det gjennomsnittlige signifikansnivået til goodness-of-fit testen når interaksjonseffekter estimeres heller.

Vi ser at når vi deler inn etter to tidgrupper og to kovariatgrupper, samt kun sampler en kontroll pr kasus blir nivået til testen høyere enn for de andre gruppene. Dette resultatet ser vi tydelig i øverst i tabell A.1 i vedlegget. Her ser vi at signifikansnivået blir høyere når det samples en enn tre kontroller for de aller fleste kombinasjonene av risikofunksjon, samplingdesign og andel sensurering. Samme tabell bekrefter regresjonsanalysens resultat om at det er en signifikant interaksjonseffekt mellom situasjonen der det i testen deles inn i åtte grupper etter kasusenes risikofunksjon og antall kontroller som samples. I dette tilfellet går effekten motsatt vei. For de aller fleste kombinasjoner av risikofunksjon, samplingdesign og andel sensurering, blir signifikansnivået til testen noe høyere når det samples tre kontroller enn når det bare samples en.

Konklusjonen etter regresjonsanalysen av simuleringsresultatene blir at selv om vi finner

noen signifikante effekter, så avviker ikke nivået noe særlig fra 5%.

For å bli ytterligere overbevist om at signifikansnivået til testen holder kunne man undersøkt dette også på andre nivåer enn 5%-nivå. Dette har vi gjort, og resultatene blir de samme som her. Disse resultatene blir derfor ikke tatt med i denne oppgaven. Istedet tar vi med QQ-plot for alle de 24 kjøringene fra situasjonene der grupperingen i testen er gjort i 2 tidgrupper og 2 kovariatgrupper. Dette plottet viser de observerte kvantilene mot de teoretiske kvantilene for hver av disse situasjonene og gir oss derfor innsikt i om testobservatorene er  $\chi^2$ -fordelt. Ingen av plottene skiller seg ut, og vi ser ingen større avvik fra den rette linjen som indikerer hvor den teoretiske og observerte kvantilen er like. QQ-plottene generaliserer dermed konklusjonene fra analysen av signifikansnivået, nemlig at dette holder også på andre nivå enn 5% nivå.

I avsnitt 3.6 presenterte vi plot av grupperte martingal residualprosesser (3.6) som en måte å evaluere hvor godt en modell passer til data. Mens goodness-of-fit testen (3.27) baserer seg på å analysere forskjellen mellom observert og estimert antall hendelser i ulike grupper, altså den grupperte martingal residualprosessen, kan den ikke avsløre hva som eventuelt er galt. Plottene kan vise mindre avvik som kan være av interesse selv om disse ikke er store nok til at testen gir forkastning, eller at den ikke gir forkastning fordi uheldig gruppering er brukt i testen.

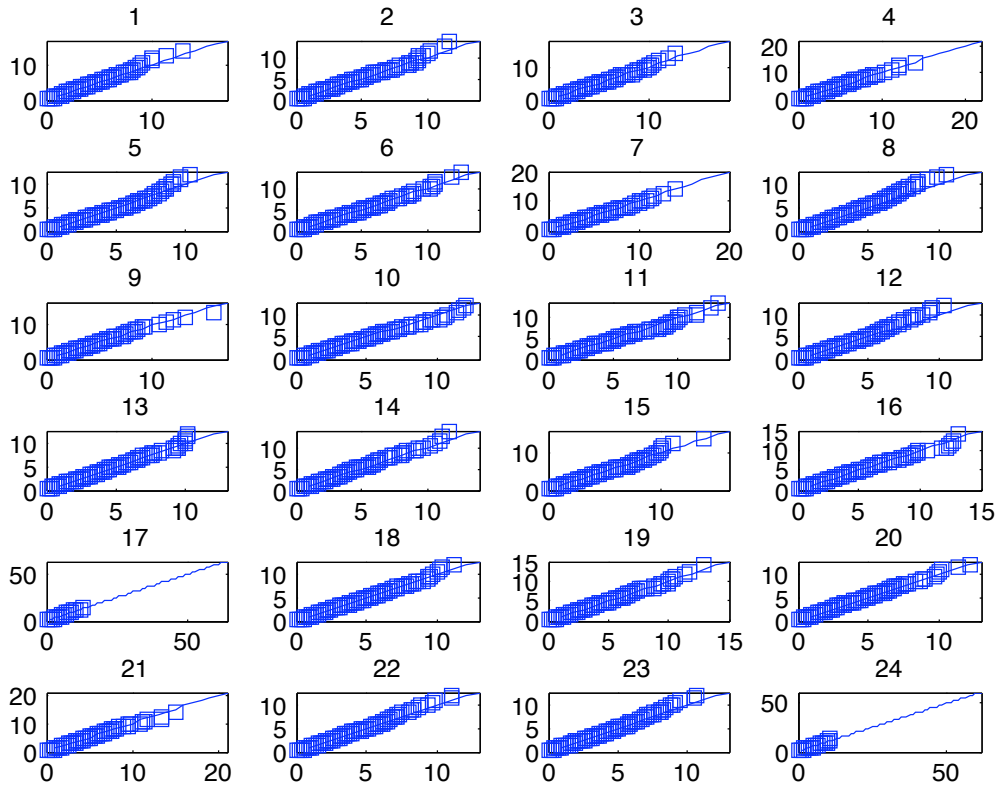
I eksempel 6 estimerte vi martingal residualprosessene som viste forskjellen på observert og estimert antall dødsfall blant urangruearbeiderne for forskjellige risikofunksjoner og samplingdesign av kontrollene. Disse er vist i figur 3.1. Individene ble delt i tre grupper etter kumulativ radoneksponering. Ingen av modellene så ut til å passe særlig bra, noe også goodness-of-fit testen (3.27) bekreftet.

På bakgrunn av plottene av de estimerte martingal residual prosessene, var vi i eksempel 6 ikke i stand til å konkludere med hva som var galt med modeltilpasningen, bare at noe ikke stemte. Ved å studere liknende plot av de estimerte martingal residual prosessene når vi simulerer, kan vi få innsikt i om disse plottene er egnet til å bestemme hva slags modellavvik vi har med å gjøre.

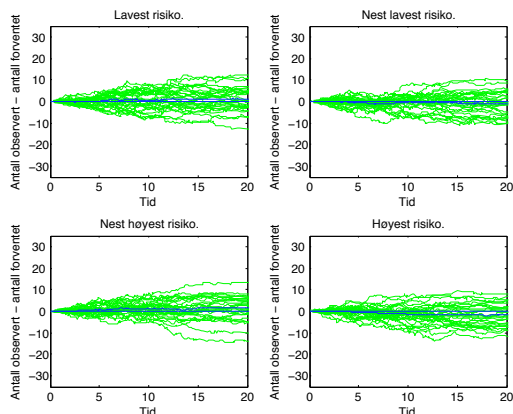
I dette avsnittet lager vi plot av de estimerte martingal residual prosessene for å undersøke om det finnes mønstre vi ikke har oppdaget i våre tidligere analyser.

Figurene 5.2 og 5.3 viser de estimerte martingal residualprosessene når modellen stemmer. Her er 80% av levetidene sensurerte, 3 kontroller er samlet ved enkel nøstet kasus kontroll til hvert kasus. Individene ble ved hver kjøring delt i fire grupper etter kvartilene til risikofunksjoenen av kasus og kontroller, og den estimerte martingal residualprosessen ble beregnet for hver gruppe.

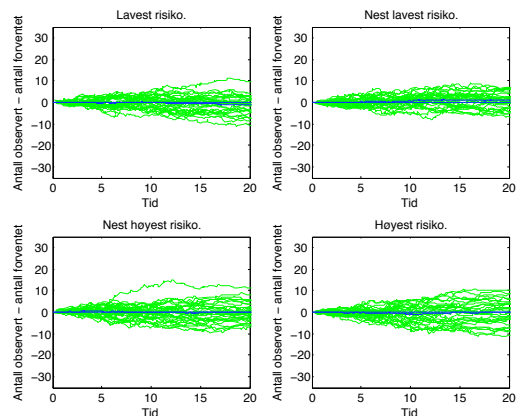
De mørke strekene som viser gjennomsnittet av de estimerte martingal residual prosessene etter 1000 kjøring, viser at det gjennomsnittlige avviket i hver av de fire gruppene er svært nær 0. De estimerte martingal residual prosessene til 30 tilfeldige kjøring viser heller ikke at det er noe systematisk avvik fra modellen i noen av tilfellene. Dette er altså slik det bør være når riktig modell er tilpasset. Ingen systematiske avvik fra modellen er synlige.



Figur 5.1: QQ-plot for alle de 24 kjøringene når modellen stemmer. Grupperingen for testen er 2 tidgrupper og 2 kovariatgrupper. I de 12 første plot er Cox-modellen brukt, mens eksess relativ risiko modellen er brukt i plottene 13 – 24. For hver risikofunksjon er de første 6 med enkel nøstet kasus kontroll, mens de neste 6 er med Counter-Matching. For hver av disse seks plottene er de to første plottene med 80% sensurering, de to neste med 90% sensurering, mens de to siste har 95% sensurering. Alle plot nummerert med partall har 3 kontroller, mens for oddetall er kun 1 kontroll trukket ved hver hendelsestid.



Figur 5.2: Estimerte martingale residualprosesser. Feil risikofunksjon tilpasset. Simulert fra og tilpasset Cox-modellen. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøring (grå) og for gjennomsnittet av 1000 kjøring (svart).



Figur 5.3: Estimerte martingale residualprosesser. Riktig modell tilpasset. Simulert fra og tilpasset eksess relativ risiko modellen. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøring (grø) og for gjennomsnittet av 1000 kjøring (svart).

## 5.5 Når modellen ikke stemmer

I dette avsnittet vil vi undersøke teststyrken til goodness-of-fit testen (3.27). Teststyrken er sannsynligheten for å forkaste nullhypotesen, at riktig modell er tilpasset, når denne hypotesen faktisk er gal. Ved å simulere levetider fra en modell og tilpasse en annen kan vi estimere teststyrken og se om den avhenger av noen av faktorene vi varierer i simuleringene våre.

Vi har valgt ut noen situasjoner vi vil undersøke teststyrken i. Disse situasjonene er realistiske i forhold til virkelige studier, og det er derfor av interesse å få undersøkt teststyrken i disse situasjonene. Vi er også interessert i å studere hvilke grupperinger for testen som er best egnet.

- Tilpasse gal risikofunksjon. Ved å simulere levetider med Cox-modellen, men tilpasse eksess relativ risiko modellen, og vice versa, kan vi undersøke om goodness-of-fit testen klarer å avsløre at feil modell er tilpasset. Dessuten får vi undersøkt om testen fungerer bedre for en av risikofunksjoene i noen av situasjonene under.
- Feil koding av kovariat. Ved å simulere levetider der en av kovariatene for eksempel er log-transformert, men tilpasse en modell hvor kovariaten forutsettes å ha en lineær effekt, får vi undersøkt om modellen avslører at feil modell er tilpasset i denne situasjonen.
- Mangler å ta med relevant kovariat. Ved å simulere levetidene med tre kovariater,

men tilpasse en modell som bare inneholder to forklaringsvariable, får vi undersøkt teststryken i denne situasjonen. Hvor stor effekt må den utelatte kovariaten ha før testen oppdager at noe mangler?

- Tidsavhengige effekter av kovariater. Begge modellene vi tilpasser forutsetter at effekten av kovariatene er konstant over tid. I virkeligheten er dette ofte ikke tilfellet. Ved å simulere levetider med tidsavhengig effekt av kovariatene og tilpasse en modell som forutsetter at effekten av kovariatene er konstante, kan vi undersøke om testen oppdager at gal modell er valgt i denne situasjonen.

### 5.5.1 Feil modell tilpasset. Liten forskjell på modellene.

I studier av overlevelsesdata er man interessert i hvilken effekt ulike forklaringsvariable har for overlevelsen til individene, og hvilken form sammenhengen mellom kovariater og overlevelse har. Gjennom valg av risikofunksjon gjør man en antakelse om hvordan et individs kovariatverdier påvirker individets hasardrate. Dersom gal modell velges vil dette kunne føre til at man misforstår hvordan forklaringsvariablene forholder seg til overlevelsen til individene. Vi er derfor interessert i å undersøke hvor godt egnet goodness-of-fit testen er til å avsløre at feil risikofunksjon er valgt. I dette avsnittet og det neste vil vi gjøre rede for teststyrken i situasjonen der gal risikofunksjon er tilpasset.

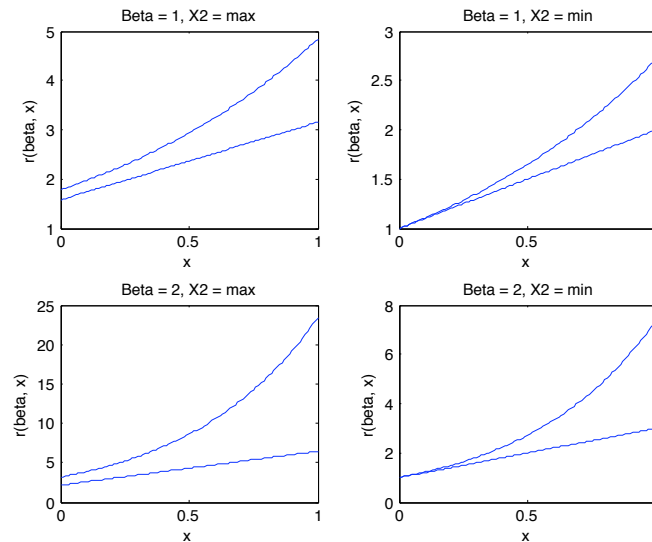
For alle de syv måtene å gruppere individene på i testen er alle de 24 kombinasjonene av risikofunksjon, samplingdesign, andel sensurering og antall kontroller gjort. Simuleringene er gjort etter samme oppskrift som i eksempel 8, med tanke på hvordan levetider er trukket og kontroller samlet, men nå er altså feil modell tilpasset. Der levetidene er simulert fra Cox-modellen er eksess relativ risiko modellen tilpasset, og motsatt.

Regresjonskoeffisientene er som i avsnitt 5.4 satt til  $\beta = (\beta_1, \beta_2)^T = (1, 1)^T$ . Dette gir, som figur 5.4 viser, liten forskjell på de to modellene vi bruker i simuleringene. I dette tilfellet tilpasser vi en krum modell (Cox) til data simulert fra en lineær modell (eksess relativ risiko), og motsatt. Det som betyr noe for "avstanden" mellom de to modellene er krummingen i Cox-modellen. Denne blir kraftigere for høyere verdier av regresjonskoeffisientene, og dermed øker forskjellen på de to modellene. I dette underavsnittet er altså modellene forholdsvis like. Vi ønsker å undersøke om testen likevel klarer å avsløre at feil modell er tilpasset.

Studier har tidligere vist at det er veldig vanskelig å skille mellom to modeller av denne typen. Formålet med våre simuleringer blir å undersøke hvordan goodness-of-fit testen (3.27) beskrevet i denne oppgaven fungerer, samt om noen av de andre faktorene vi varierer, som andel sensurering eller gruppering i testen, har effekter på teststyrken i en slik situasjon.

Simuleringsresultatene er gjengitt i tabell A.4 i vedlegget. Siden de to modellene er såpass like i denne situasjonen, kommer det ikke overaskende at teststyrken er så lav som man får inntrykk av her. Tabell 5.7 viser resultatene av regresjonsanalyse uten interaksjonseffekter av simuleringsresultatene når feil modell er tilpasset, men det er liten forskjell på modellene. Regresjonsanalyse av simuleringsresultatene bekrefter inntrykket. Den gjennom-





Figur 5.4: Riskofunksjonene til Cox og eksess relativ risiko -modellen for ulike verdier av regresjonskoeffisientene. De to øverste figurene viser risikofunksjonene når  $\beta = (1, 1)^T$ . De to nederste figurene viser risikofunksjonene når  $\beta = (2, 2)^T$ . Risikofunksjonen til Cox-modellen er kurvet, mens risikofunksjonen til eksess relativ risiko modellen er lineær.

snittlige teststyrken blir av analysen uten interaksjonseffekter estimert til 5.4%. I avsnitt 5.4.1 estimerte vi signifikansnivået til 5.3% når interaksjoner ble utelatt. Regresjonsanalysen med signifikante interaksjonseffekter anslår teststyrken til 5.1%. Teststyrken er altså svært lav i denne situasjonen.

Regresjonsanalysen uten samspilleffekter viser at de to samme hovedeffektene blir signifikante på 1% nivå nå som når modellen stemmer. Disse er grupperingen der individene deles i åtte grupper etter kasusenes risikofunksjoner og antall samplete kontroller. Størrelsen på disse effektene er også tilnærmet de samme. Vi kan derfor ikke konkludere med at testen har høyere teststyrke i situasjonen der det i testen deles inn i 8 grupper etter kasusenes risiko, eller at den har lavere styrke når kun 1 kontroll samples. I disse situasjonene er det nivået som ikke holder.

Resultatene av regresjonsanalyse med alle signifikante 1. ordens interaksjoner er gitt i tabell A.3 i vedlegget. Vi ser at det i likhet med da vi tilpasset riktig modell, kun er interaksjoner mellom hvilken gruppering som brukes i testen og antall kontroller som blir signifikante på 1% nivå. Nå som sist er dette også de to samme effektene, og de er nesten nøyaktig like store.

Figur 5.5 viser de estimerte martingal residual prosessene fra 30 tilfeldige kjøring (grå) og gjennomsnittet av 1000 kjøring (svart). Her er levetidene simulert fra Cox-modellen, mens eksess relativ risiko modellen er tilpasset og det er liten forskjell på modellene ( $\beta_1 = \beta_2 = 1$ ). I hver kjøring er 3 kontroller trukket ved hver hendelsestid ved enkel nøstet

Tabell 5.7: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Feil modell tilpasset. Liten forskjell på modellene. Koefisienten for Cox betyr at Cox-modellen er tilpasset, det er med andre ord simulert fra eksess relativ risiko modellen.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Intercept	-2.860	0.011	-262.629	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.039	0.027	-1.453	0.146
4RiskR	0.036	0.026	1.354	0.176
8RiskR	-0.022	0.027	-0.808	0.419
4RiskKasus	-0.022	0.027	-0.808	0.419
8RiskKasus	0.132	0.025	5.198	$2.01 \cdot 10^{-7}$
2Tid3RiskR	-0.002	0.027	-0.070	0.945
Cox	0.025	0.011	2.288	0.022
80%c	-0.015	0.015	-0.996	0.319
90%c	0.019	0.015	1.217	0.224
KasusKontroll	0.000	0.011	0.026	0.979
1Kontroll	-0.073	0.011	-6.744	$1.54 \cdot 10^{-11}$

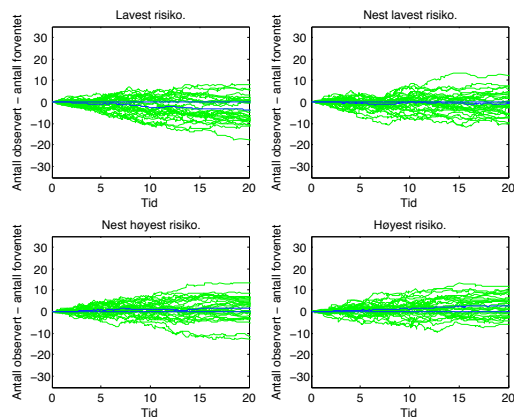
kasus kontroll. Figur 5.6 viser det samme, men her er levetidene simulert fra eksess relativ risiko, mens Cox-modellen er tilpasset. Individene er gruppert etter kasus og kontrollers risikofunksjoner. Figurene bekrefter inntrykket fra den øvrige analysen. Vi ser ingen klare mønstre i de estimerte martingal residualene som tyder på at feil modell er tilpasset.

I dette avsnittet har vi sett at testens evne til å avsløre at feil modell er tilpasset er svært liten når den tilpassete modellen er så lik den sanne som i dette tilfellet. I neste avsnitt vil vi øke forskjellen på modellen vi simulerer fra og den vi tilpasser og se hvordan dette påvirker teststyrken. Vi er også interessert i om dette får konsekvenser for hvilke av faktorene som blir signifikante.

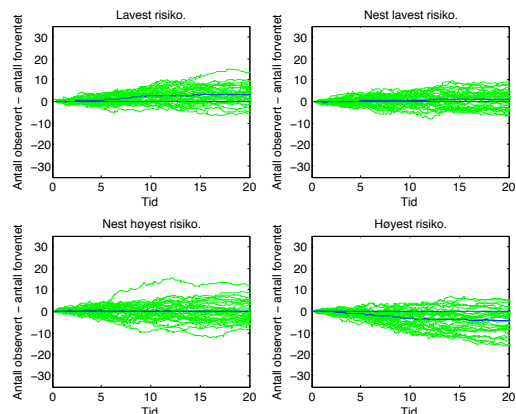
### 5.5.2 Feil modell tilpasset. Stor forskjell på modellene.

I dette avsnittet vil vi øke forskjellen på modellen vi simulerer levetider fra og modellen vi tilpasser til data, for å se om teststyrken øker eller det blir endringer i signifikante faktorer. Plottene nederst i figur 5.4 viser forskjellen på risikofunksjon til Cox-modellen og eksess relativ risiko modellen når begge regresjonskoefisientene settes til 2 i stedet for 1 som i avsnitt 5.5.1. Krummingen til Cox-modellen er kraftigere nå enn i avsnitt 5.5.1, og dermed er forskjellen på de to modellene større. I dette avsnittet vil vi undersøke styrken til goodness-of-fit testen (3.27) i denne situasjonen.

Andel forkastninger for hver av de syv mulige grupperingene for testen er gitt i tabell A.7 i vedlegget. ER betyr her at det er simulert fra eksess relativ risiko modellen, mens Cox-modellen er tilpasset. COX betyr det motsatte. Tabellen viser at teststyrken ikke er særlig høy heller i denne situasjonen, men også at styrken varierer mellom situasjonene.



Figur 5.5: Estimerte martingal residualprosesser. Feil risikofunksjon tilpasset. Simulert fra eksess relativ risiko modellen, tilpasset eksess relativ risiko modellen, og det er liten forskjell på modellene. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingal residualprosessene for 30 tilfeldige kjøringer (grå) og for gjennomsnittet av 1000 kjøringer (svart).



Figur 5.6: Estimerte martingal residualprosesser. Feil risikofunksjon tilpasset. Simulert fra eksess relativ risiko modellen, tilpasset Cox. Liten forskjell på modellene. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingal residualprosessene for 30 tilfeldige kjøringer (grå) og for gjennomsnittet av 1000 kjøringer (svart).

Først og fremst er det en tydelig forskjell på modellene. Når levetidene simuleres fra Cox-modellen og eksess relativ risiko modellen tilpasses, er teststyrken langt høyere enn i det motsatte tilfellet. Videre ser det ut til å være en klar sammenheng mellom andel sensurerte levetider og teststyrke. Mindre andel sensur gir høyere teststyrke. Når sensurandelen er lav får vi flere kasus og dermed mer informasjon å både tilpasse modellen på og gjøre testen på. Å avsløre forskjellen mellom to modeller blir dermed enklere med mindre sensureringsandel.

Mer informasjon får vi også hvis vi trekker 3 kontroller enn hvis vi bare trekker 1 kontroll per kasus. I tabellen ser det ut til at tre kontroller gir noe høyere teststyrke enn når kun en kontroll samples. Spesielt ser det ut til at teststyrken blir lavere når kun en kontroll samples ved enkel nøstet kasus kontroll.

Resultatene av logistisk regresjonsanalyse uten interaksjonseffekter av simuleringsresultatene er gitt i tabell 5.8. Her estimeres gjennomsnittlig teststyrke til 6.6%. Nesten alle de syv grupperingene blir nå signifikante på 1% nivå. Den estimerte effekten av grupperingen 2Tid4RiskR blir  $-0.1326$ . Unntaket er når vi bare deler inn i fire grupper etter kasus eller kasus og kontrollers risiko. Når tid brukes i grupperingen blir nå grupperingen alltid signifikant. Vi ser at de to siste situasjonene, der grupperingen gjøres etter tid og risiko til kasus og kontroller, har en negativ effekt på styrken. De andre situasjonene har en positiv estimert effekt på styrken, men i to av tilfellene er denne altså ikke signifikant. Høyest teststyrke får situasjonene der individene deles inn i åtte grupper kun etter risiko

Tabell 5.8: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Feil modell tilpasset. Stor forskjell på modellene.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Intercept	-2.651	0.010	-257.393	$< 2 \cdot 10^{-16}$
2Tid2Kov	0.080	0.024	3.405	0.001
4RiskR	0.036	0.024	1.471	0.141
8RiskR	0.076	0.024	3.173	0.002
4RiskKasus	0.019	0.025	0.772	0.440
8RiskKasus	0.070	0.024	2.942	0.003
2Tid2RiskR	-0.148	0.026	-5.776	$7.66 \cdot 10^{-9}$
Cox	0.228	0.010	22.589	$< 2 \cdot 10^{-16}$
80%c	0.195	0.014	14.417	$< 2 \cdot 10^{-16}$
90%c	-0.036	0.014	-2.500	0.012
KasusKontroll	-0.034	0.010	-3.366	0.001
1Kontroll	-0.094	0.010	-9.407	$< 2 \cdot 10^{-16}$

og den første situasjonen med gruppering etter tid og kovariat.

Vi får høyere teststyrke når vi simulerer fra Cox-modellen og tilpasser eksess relativ risiko modellen. Den gjennomsnittlige teststyrken blir i denne situasjonen estimert til 8.1%. I det motsatte tilfellet, der levetidene simuleres fra eksess relativ risiko modellen mens Cox-modellen tilpasses, estimeres teststyrken til 5.3%.

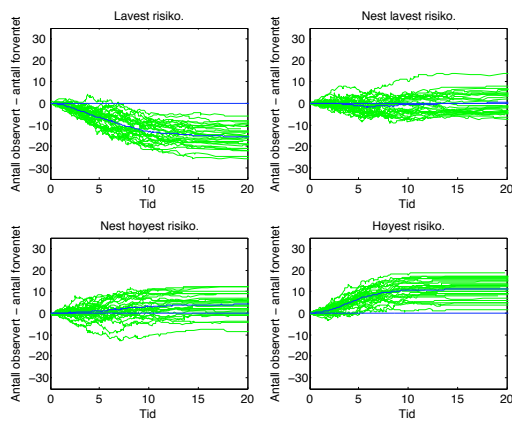
At andel sensurerte levetider har signifikant effekt på teststyrken, bekreftes av regresjonsanalysen. Når 80% av levetidene sensureres blir teststyrken estimert til 7.9%, mens den blir estimert til lavere enn den gjennomsnittlige teststyrken for de to lavere andelene sensurering. For 90% er denne effekten ikke signifikant på 1% nivå.

Effekten av samplingdesign og antall kontroller er begge signifikante, men små. Ikke uventet gir enkel nøstet kasus-kontroll noe lavere teststyrke enn Counter-Matching, mens 1 kontroll gir noe lavere teststyrke enn 3 kontroller.

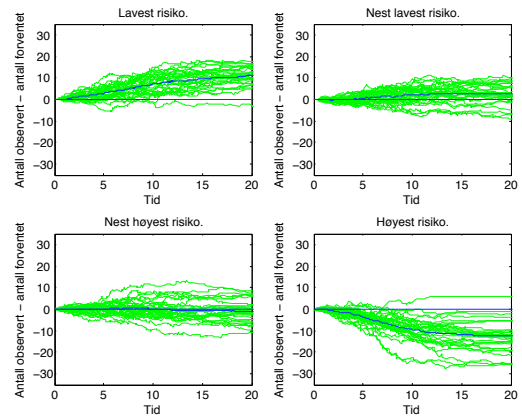
Når interaksjoner inkluderes i regresjonsmodellen, estimeres den gjennomsnittlige teststyrken til 6.1%. Effekten av de syv grupperingene i testen blir mindre signifikante enn før interaksjoner ble tatt inn i modellen. Nå er det kun de to siste grupperingene, der det grupperes etter tid og risiko til kasus og kontroller, som blir signifikante på 1% nivå. Risikofunksjon og andel sensur er fortsatt signifikante faktorer, mens samplingdesign ikke lenger blir signifiant. Det blir derimot fortsatt antall samplete kontroller, og effekten er her omtrent den samme som da interaksjoner ble utelatt.

Interaksjonseffektene mellom de fleste grupperingene og risikofunksjon blir signifikante, men effektene er små. Det er to signifikante interaksjonseffekter mellom gruppering og andel sensurerte levetider. Den ene er mellom grupperingen 8RiskR og 80% sensurering. Her blir teststyrken 8.1%, mens den andre er mellom 2Tid2RiskR og 90% sensurering. Her blir den estimerte teststyrken så lav som 3.2%.

Mellom gruppering og antall kontroller er det bare ett signifikant samspill. Dette er mellom grupperingen 8RiskKasus og 1 kontroll. Når disse to opptrer samtidig estimeres styrken til noe lavere enn dersom 3 kontroller trekkes i samme situasjon.



Figur 5.7: Estimerte martingale residualprosesser. Feil risikofunksjon tilpasset. Simulert fra Cox-modellen, tilpasset eksess relativ risiko modellen. Stor forskjell på modellene. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøringer (grå) og for gjennomsnittet av 1000 kjøringer (svart).



Figur 5.8: Estimerte martingale residualprosesser. Feil risikofunksjon tilpasset. Simulert fra eksess relativ risiko modellen, tilpasset Cox. Stor forskjell på modellene. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøringer (grå) og for gjennomsnittet av 1000 kjøringer (svart).

Figur 5.7 og 5.8 viser de estimerte grupperte martingale residual prosessene, når individene deles i fire grupper etter kvartilene til kasusenes risikofunksjoner. De grå strekene er de estimerte martingale residual prosessene fra 30 tilfeldige kjøringer, mens den svarte streken er gjennomsnittet av 1000 kjøringer. I figur 5.7 er levetidene simulert fra Cox-modellen, mens eksess relativ modellen er tilpasset. I figur 5.8 er det motsatt. I begge situasjonene er andelen sensurerte levetider 80% og tre kontroller er trukket ved hver hendelsestid.

Plottene illustrerer hvordan modellmisstilpasningen, særlig i gruppene med høyest og lavest risiko, fører til relativt store avvik mellom observert og forventet antall dødsfall. Vi legger merke til at fortegnet på avvikene er motsatt. Selv om avviket mellom observert og forventet antall dødsfall for noen grupper kan nærme seg 20, er altså ikke dette nok til at goodness-of-fit testen avslører at feil modell er valgt.

Konklusjonen etter simuleringene der feil risikofunksjon er tilpasset blir at goodness-of-fit testen bare i liten grad egnert seg til å avsløre at feil modell er tilpasset, selv når modellene er såpass ulike som i dette tilfellet. Videre har vi sett hvordan teststyrken blir høyere når vi tilpasser eksess relativ risiko modellen til data simulert fra Cox semiparametriske regresjonsmodell enn omvendt. Teststyrken avhenger også av andel sensur, men selv ikke når andelen er lav blir teststyrken høy.

### 5.5.3 Tidsavhengige effekt av kovariat. Liten positiv effekt.

En realistisk situasjonen i virkelige medisinske studier er at en kovariat endrer effekt over tid. En modell med konstant effekt av kovariaten vil da kunne føre til systematisk gal estimering av effekten til kovariaten. Modellen vil estimere et gjennomsnitt av den tidsavhengige effekten. Dersom effekten av kovariaten avtar over tid, vil dette konstante estimerte nivået være lavere enn den egentlige effekten til å begynne med, mens det vil bli høyere enn den egentlige effekten når denne har avtatt. Dette vil føre til at modellen predikerer for få hendelser til å begynne med og for mange senere. Hvis den tidsavhengige effekten av kovariaten er økende blir situasjonen motsatt. Dersom vi tilpasser en modell der effekten av kovariaten antas å være konstant over tid, vil dette dermed gi en gal forståelse av effekten til kovariaten. Vi er derfor interessert i å avsløre om feil modell er valgt. I dette avsnittet vil vi gjøre rede for styrken til goodness-of-fit testen (3.27) når vi tilpasser Cox semiparametriske regresjonsmodell til data hvor en kovariat har en effekt som endrer seg over tid.

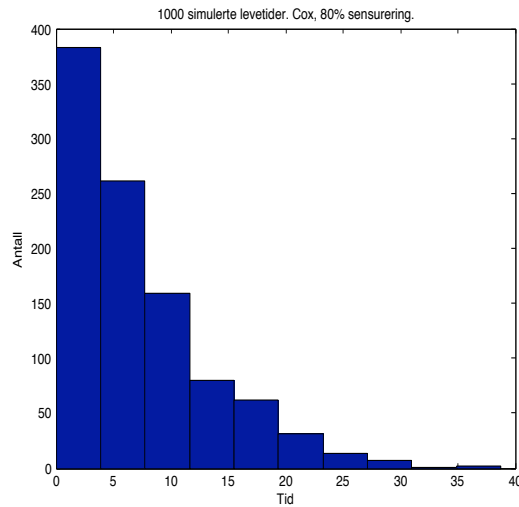
Teorien bak simulering av levetider der effekten av en kovariat er tidsavhengig, er beskrevet i avsnitt 4.3.1. Data simuleres i utgangspunktet som i eksempel 8, men effekten av den uniforme kovariaten  $X_1$  gjøres nå tidsavhengig. Levetidene  $T$  finnes nå ved (4.12).

Størrelsen på tidsavhengigheten er avgjørende for hvor stor forskjell det er på den estimerte konstante effekten av kovariaten og den egentlige effekten av forklaringsvariabelen ved en gitt tid. Når vi trekker levetider som i (4.12), kan vi justere graden av tidsavhengighet ved å justere parameteren  $c$ . I dette avsnittet vil vi undersøke teststyrken for to ulike nivåer av tidsavhengighet. Effekten av kovariaten over tid er vist ved blå linje i figur 5.10. I begge tilfellene lar vi effekten av den uniformt fordelte variabelen  $X_1$  være tidsavhengig, mens effekten av den diktome variabelen  $X_2$  forblir konstant. Det eneste som endres i simuleringene i forhold til situasjonen i eksempel 8 er dermed at effekten av  $X_1$  endres over tid. Cox-modellen brukes både i simuleringene og tilpasses i hvert tilfelle.

Instillingen av tidsavhengigheten ble gjort både ut fra tanker om hva som kunne være realistiske situasjoner, samt noe prøving og feiling. Den første situasjonen vi vil simulere er en forlengelse av standardsituasjonen presentert i eksempel 8. Vi lar, som tidligere, effekten av  $X_1$  begynne med verdi 1. Så lar vi  $\beta_1(t)$  øke slik at den er omtrent firedoblet ved tid 10. Figur 5.9 viser at dette er noe etter gjennomsnittlig levetid ved 80% sensurering. Dette oppnår vi ved å sette verdien til parameteren  $c$  i (4.12) til 3. Verdien til  $\beta_1(t)$  i dette tilfellet er illustrert med blå linje i figur 5.10. De estimerte konstante regresjonskoeffisientene for de første ti kjøringene er vist som rette linjer i figur 5.11. Den krumme linjen viser  $\beta_1(t)$  brukt i simuleringene.

De estimerte konstante regresjonskoeffisientene i figur 5.11 ligger alle mellom 0.73 og 2.04. Vi ser at effekten av  $X_1$  i de fleste tilfellene estimeres til et nivå som ligger noe over startnivået til  $\beta_1(t)$ . To av de estimerte konstante regresjonskoeffisientene ligger på et nivå under 1, som er startnivået til  $\beta_1(t)$ .

Tabell A.9 i vedlegget viser at teststyrken er lav. Den høyeste teststyrken oppnås ikke uventet når tid er en faktor i grupperingen til testen. I disse situasjonene deles individene

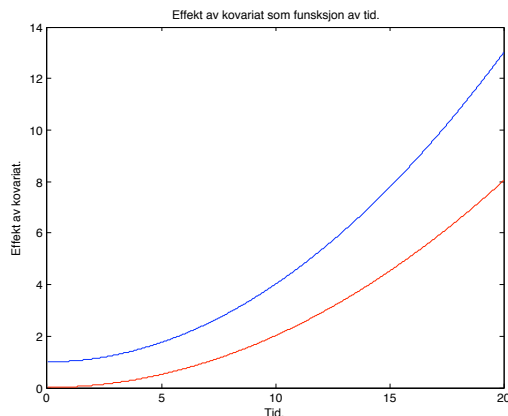


Figur 5.9: Histogram over 1000 simulerte sensurerte levetider fra Cox-modellen. Her er  $\beta_1 = \beta_2 = 1$ . Andelen sensurering er 80%.

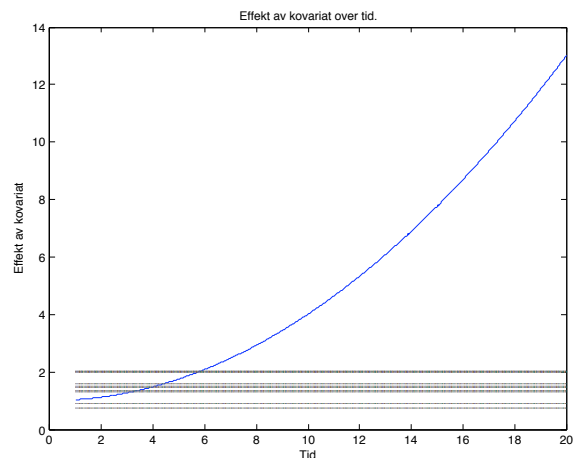
inn i to tidgrupper. Siden effekten av  $X_1$  øker med tiden vil effekten av denne gjerne være noe høyere i den siste tidgruppen enn den estimerte konstante effekten av kovariaten. I første tidgruppe blir det motsatt. Derfor er det naturlig at testen har høyere styrke når tid brukes som en faktor i grupperingen til testen. Når tidsavhengigheten ikke er kraftigere enn i denne situasjonen blir styrken likevel ikke særlig høy selv om grupperingen gjøres etter tid.

Resultatene av regresjonsanalyse av simuleringresultatene uten interaksjonseffekter er gitt i tabell 5.9. Regresjonsanalysen av simuleringresultatene bekrefter at den estimerte gjennomsnittlige teststyrken blir noe høyere når tid er en av faktorene i grupperingen til testen, men heller ikke her blir styrken høy. Den estimerte gjennomsnittlige teststyrken blir på 6.0%. Grupperingen som viser seg å gi størst teststyrke er grupperingen etter tid og kovariat. Den gjennomsnittlige teststyrken blir i denne situasjonen estimert til 7.3%. Andel sensur har også signifikant effekt på signifikansnivået, men størrelsen på denne effekten er forholdsvis liten. Hverken sampledesign og antall samplete kontroller blir signifikante på 1% nivå.

Konklusjonen blir at testen har liten evne til å avsløre at feil modell er tilpasset når den tidsavhengige effekten ikke er kraftigere enn i dette tilfellet. De observerte testobservatorene er svært nære fordelt som nullfordelingen. For å avsløre tidsavhengige effekter ser det ut til at tid bør inngå som en faktor i grupperingen til testen. Da deles individene inn i to tidgrupper. Siden effekten av  $X_1$  øker med tiden, vil effekten av denne gjerne være noe høyere i den siste tidgruppen enn den estimerte konstante effekten av kovariaten. I første tidgruppe blir det motsatt. Derfor er det naturlig at testen har høyere styrke når tid brukes som en faktor i grupperingen til testen. Når tidsavhengigheten ikke er kraftigere enn i denne situasjonen blir styrken likevel ikke særlig høy selv om grupperingen gjøres etter tid. I neste avsnitt vil vi øke den tidsavhengige effekten av kovariaten noe og se om



Figur 5.10: Plot av de to ulike tidsavhengige effektene av  $X_1$  fra avsnitt 5.5.3. I det første tilfellet begynner  $\beta_1(t)$  med verdi 1, for så å vokse til ca 4 ved tid 10 (blå), som er noe etter gjennomsnittlig levetid ved 80% sensurering. I det andre tilfellet begynner  $\beta_1(t)$  med verdien 0.01 for så å vokse til omtrent 2 ved tid 10 (rød).



Figur 5.11: Plot av  $\beta_1(t)$  når vi har en liten positiv tidsavhengig effekt av en kovariat. De vannrette linjene viser den estimerte konstante effekten  $\hat{\beta}$  av kovariaten fra de ti første kjøringene. Den kurvete linjen viser den tidsavhengige  $\beta_1(t)$  brukt i simuleringene.

teststyrken øker da.

#### 5.5.4 Tidsavhengig effekt av kovariat. Stor positiv effekt.

Som vi så i forrige avsnitt ble teststyrken for lave verdier av  $c$  svært lav. I dette avsnittet ser vi på en situasjon der  $\beta_1(t)$  har en startverdi på 0.01, mens  $c$  ble valgt slik at denne vokste til omkring 2 ved tid 10. Dette er noe etter gjennomsnittlig verdi av levetidene i tilfellet med 80% sensurering. Parameteren  $c$  må da ta verdien 200. Den røde linjen i figur 5.11 illustrerer hvordan  $\beta_1(t)$  endres over tid.

Risikofunksjon gikk i forrige avsnitt for et individ med maksimale kovariatverdier av  $X_1$  fra  $e^1 = 2.7$  ved tid 0 til  $e^4 = 54.6$  ved tid 10. Dette betyr at risikofunksjonen ble omtrent 20-doblet over dette tidsintervallet. I dette avsnittet lar vi risikofunksjonen for et individ med maksimal verdi av  $X_1$  gå fra  $e^{0.01} = 1.0$  til  $e^2 = 7.4$  i samme tidsrom. Dette svarer til noe over en syvdobling av risikofunksjonen.

Tabell A.11 i vedlegget oppsummerer resultatene av simuleringene. Som vi ser av tabellen er teststyrken nå, i de tilfellene tid er en faktor i grupperingen til testen, svært mye høyere enn den har vært tidligere. Når 80% av levetidene sensureres, 3 kontroller samples til hvert kasus ved Counter-Matching og grupperingen i testen gjøres etter tid og kovariat, blir teststyrken på 89%. Da vi hadde mindre tidsavhengighet i forrige avsnitt var teststyrken for samme situasjon 9%. Også når tid inngår i grupperingen til testen sammen med risiko, blir teststyrken svært mye høyere enn tidligere. Teststyrken blir noe høyere



Tabell 5.9: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Tidsavhengige effekt av kovariat. Liten positiv effekt.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.760	0.015	-187.645	$< 2 \cdot 10^{-16}$
2Tid2Kov	0.222	0.033	6.718	$1.85 \cdot 10^{-11}$
4RiskR	-0.121	0.037	-3.224	0.001
8RiskR	-0.146	0.038	-3.858	0.000
4RiskKasus	-0.090	0.037	-2.442	0.015
8RiskKasus	0.065	0.035	1.864	0.062
2Tid2RiskR	0.095	0.035	2.747	0.006
80%c	0.129	0.020	6.424	$1.32 \cdot 10^{-10}$
90%c	-0.058	0.021	-2.772	0.006
KasusKontroll	-0.031	0.015	-2.162	0.031
1kontroll	-0.026	0.015	-1.817	0.069

når risikogrupperingen gjøres ved å dele kasus og kontrollers risikofunksjoner i to enn i fire grupper. I situasjonene der grupperingen til testen gjøres etter kun risiko, blir teststyrken noe høyere enn tidligere, men kommer bare i to situasjoner over 10%.

Resultatene av regresjonsanalysen av simuleringsresultatene uten interaksjonseffekter er gitt i tabell 5.10. Alle faktorene er nå signifikante på 1% nivå. Den gjennomsnittlige teststyrken estimeres til 14.7%. Størrelsen på effektene er nå store og det er klare forskjeller på de ulike grupperingene. Inntrykket av at teststyrken øker når tid inngår i grupperingen til testen, bekreftes. Alle de tre grupperingene hvor tid inngår har positive effekter, det vil si at teststyrken øker. Høyest teststyrke får vi også nå ved å gruppere etter tid og kovariat. I dette tilfellet blir den estimerte gjennomsnittlige teststyrken 48.7%. Lavest blir den estimerte teststyrken når individene deles i fire grupper etter kasusenes risikofunksjoner.

Videre ser vi at effekten av sensurering øker kraftig i forhold til forrige avsnitt. Den gjennomsnittlige teststyrken ved 80% sensurering estimeres nå til 25.6%. For 90% og 95% sensurering er de samme estimatene henholdsvis 13.8% og 8.4%. Også effekten av samplingdesign og antall samplate kontroller blir nå større enn tidligere, men har mindre å si for teststyrken enn gruppering for testen eller andel sensurering.

Resultatene av regresjonsanalysen av simuleringsresultatene med signifikante 1. ordens samspilleffekter, er gitt i tabell A.10 i vedlegget. Samspilleffektene mellom andel sensurering og samplingdesign eller antall kontroller er ikke lenger signifikante på 1% nivå. Gjennomsnittlig teststyrke estimeres nå til 12.2%. Hovedeffektene av gruppering blir noe lavere enn da vi ikke tok med samspilleffekter.

Det er signifikante samspilleffekter mellom gruppering ved testen og andel sensur. I de situasjonene hvor tid brukes i grupperingen blir den estimerte gjennomsnittlige teststyrken høyere når 80% sensurering velges, enn for høyere andeler sensurering. Dette er de samme grupperingene som i utgangspunktet har en positiv hovedeffekt. For eksempel estimeres den gjennomsnittlige teststyrken i situasjonen med gruppering i to tidgrupper og

Tabell 5.10: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Tidsavhengige effekt av kovariat. Stor positiv effekt.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-1.761	0.012	-146.354	$< 2 \cdot 10^{-16}$
2Tid2Kov	1.708	0.020	84.550	$< 2 \cdot 10^{-16}$
4RiskR	-1.016	0.033	-30.964	$< 2 \cdot 10^{-16}$
8RiskR	-0.963	0.032	-29.903	$< 2 \cdot 10^{-16}$
4RiskKasus	-1.090	0.034	-32.343	$< 2 \cdot 10^{-16}$
8RiskKasus	-0.980	0.032	-30.254	$< 2 \cdot 10^{-16}$
2Tid2RiskR	1.308	0.020	64.464	$< 2 \cdot 10^{-16}$
80%c	0.695	0.013	52.197	$< 2 \cdot 10^{-16}$
90%c	-0.068	0.014	-4.979	$6.38 \cdot 10^{-7}$
KasusKontroll	-0.376	0.010	-38.395	$< 2 \cdot 10^{-16}$
1Kontroll	-0.139	0.010	-14.390	$< 2 \cdot 10^{-16}$

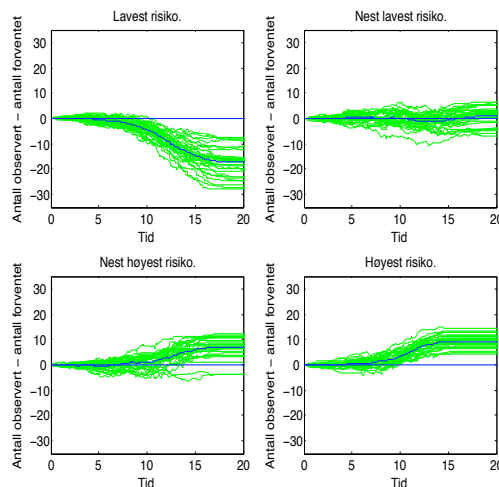
to kovariatgrupper og 80% sensurering til 27.8% Når grupperingen i testen gjøres kun etter risiko er det derimot estimert en negativ effekt av 80% sensurering, i forhold til det gjennomsnittlige estimerte signifikansnivået.

Det er noen samspilleffekt mellom gruppering i testen og samplingdesign som er signifikant på 1% nivå. Disse effektene er små. Også mellom gruppering i testen og antall kontroller er det små, men signifikante interaksjonseffekter. Her er det interessant å merke seg retningen på effektene. Når tid inngår i grupperingen til testen, blir den estimerte teststyrken høyere ved å sample 3 kontroller enn kun 1. Når grupperingen kun gjøres etter risiko, blir derimot den estimerte teststyrken noe lavere når 3 kontroller samples enn når kun 1 kontroll samples ved hvert hendelsestidspunkt.

Til slutt er det en svært signifikant samspilleffekt mellom samplingdesign og antall kontroller. Dette er også synlig fra tabell A.11. Her ser vi at når kontrollene samples ved Counter-Matching blir teststyrken oftere høyest når kun 1 kontroll samples enn når 3 kontroller samples. Ved Enkel nøstet kasus kontroll er effekten motsatt. Denne høyeste styrken ved enkel nøstet kasus kontroll når tre kontroller samples, blir derimot sjelden like høy eller høyere enn styrken ved Counter-Matching.

Figur 5.12 viser de estimerte grupperte martingal residualprosessene. Individene er delt i fire grupper basert på deres individuelle estimerte relative risikofunksjon. Plottene viser hvordan modellen, som ikke tar hensyn til at effekten av en kovariat er tidsavhengig, men tilpasser et gjennomsnittlig konstant nivå for effekten av kovariaten, avviker fra virkeligheten. For individene med lavest estimert individuell risikofunksjon (øverst til venstre), estimeres for mange dødsfall, mens for de to gruppene med høyest estimert risikofunksjon (nederst), estimeres modellen for få dødsfall. For gruppen med nest lavest risiko (øverst til høyre), passer modellen ganske godt, og vi ser ingen mønstre i martingal residual prosessene.

Det er interessant å legge merke til at martingal residual prosessene estimert fra situasjonen



Figur 5.12: Estimerte martingale residualprosesser. Stor tidsavhengig effekt av kovariat. Andelen sensurerte levetider er 80%, 3 kontroller er samplet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøring (grå) og for gjennomsnittet av 1000 kjøring (svart).

med tidsavhengig effekt av kovariat, likner mye på de i figur 5.7. Her tilpasset vi feil modell, men hadde ingen tidsavhengig effekt av kovariat. For det første ser vi hvordan de gjennomsnittlige martingale residual prosessene i figur 5.12 ligger rundt 0 frem til noe før tid 10. Før dette tidspunktet er forskjellen på tilpasset modell og den som er brukt i simuleringene, svært små siden den tidsavhengige effekten av  $X_1$  kun i liten grad har begynt å gjøre seg gjeldene. Etter dette tidspunktet vokser avviket og ender på omtrent samme nivå som i avsnitt 5.5.2.

Selv om avviket mellom antall observerte og estimerte dødsfall er omtrent det samme i de to situasjonene, ender vi altså med svært ulik teststyrke. Den høye teststyrken i dette avsnittet oppnådde vi først da grupperingen for testen ble gjort etter tid. Når grupperingen ble gjort etter risiko er teststyrken i de to situasjonene ikke særlig ulik. I figurene 5.12 og 5.8 er martingale residual prosessene gruppert etter risikofunksjonen til kasus og kontroller. Dette tilsvarer når vi i testen bruker grupperingen 4RiskR. Testen ser da på avviket mellom observert og forventet antall dødsfall ved endetidspunktet og ender med omtrent samme resultat for de to situasjonene. Når grupperingen også gjøres etter tid, blir det lettere for testen å avsløre forskjellen på de to modellene. Dette forklarer forskjellen i teststyrke i de to situasjonene.

Hovedkonklusjonen etter simuleringene blir at styrken til goodness-of-fit testen avhenger svært mye av hvordan grupperingen i testen gjøres. Dersom individene deles inn etter tid i grupperingen til testen, blir teststyrken svært mye høyere enn hvis grupperingen kun gjøres etter risiko. Viktige resultater ellers er at styrken blir høyere med lavere andel sensurerte levetider. Dette gjelder særlig når grupperingen gjøres etter tid. Sampledesign og antall kontroller som samples ved hver hendelsestid påvirker også teststyrken, men i mindre

grad. Goodness-of-fit testen (3.27) viser seg ganske effektiv for å oppdage modellavvik i situasjonen med tidsavhengig effekt av kovariat, men tidsavhengigheten må være på et forholdsvis høyt nivå for at testen skal avsløre denne.

### 5.5.5 Manglende kovariat.

I virkeligheten har man sjelden oversikt over alle variable av betydning. I virkelige kasus-kontroll studier er dette heller ikke tilfellet. Vi kan ikke regne med å ha full oversikt over alle variable av betydning, men dersom viktige forklaringsvariable er utelatt fra modellen vår er vi interessert i å oppdage dette. I dette avsnittet vil vi gjøre rede for styrken til goodness-of-fit testen (3.27) når vi tilpasser en modell som mangler en kovariat.

Ved å innføre en tredje kovariat som vi bruker i simuleringene, men ikke når vi tilpasser modellen, kan vi simulere denne situasjonen. Den tredje kovariaten  $X_3$  lar vi være uniformt fordelt over  $(0, 1)$ . Vi trekker nå levetiden  $T$  som i (4.1), men risikofunksjonen er nå gitt ved  $r(\beta, X) = e^{\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$ . Ved å variere  $\beta_3$  kan vi undersøke hvor viktig den manglende kovariaten må være før goodness-of-fit testen skal oppdage at feil modell er tilpasset.

Vi har simulert to ulike situasjoner der en kovariat mangler i den tilpassete modellen. I det første tilfellet er effekten av den manglende kovariaten like stor som de som er inkludert i den tilpassete modellen. Her er altså  $\beta = (1, 1, 1)^T$ . I den andre situasjonen lar vi den manglende kovariaten ha høyere effekt enn de utelatte. I dette tilfellet lar vi  $\beta_3 = 2$ . De to andre kovariatene har fortsatt samme effekt. Cox-modellen ble brukt både i simuleringene og tilpasset.

Simuleringsresultatene fra den første situasjonen er gitt i tabell A.13 i vedlegget. Simuleringsresultatene når  $\beta_3 = 2$  er gitt i tabell A.15 i vedlegget. Vi ser at teststyrken er svært lav i begge situasjoner.

Resultatene av en regresjonsanalyse av simuleringsresultatene er gitt i tabell 5.11, for situasjonen der  $\beta_3 = 1$ , og i tabell 5.12 når  $\beta_3 = 2$ . Den gjennomsnittlige teststyrken estimeres i de to situasjonene til 5.2% og 5.3% henholdsvis. For begge situasjonene gjelder det at kun antall kontroller har signifikant effekt på teststyrken. Størrelsen på denne effekten er imidlertid svært liten.

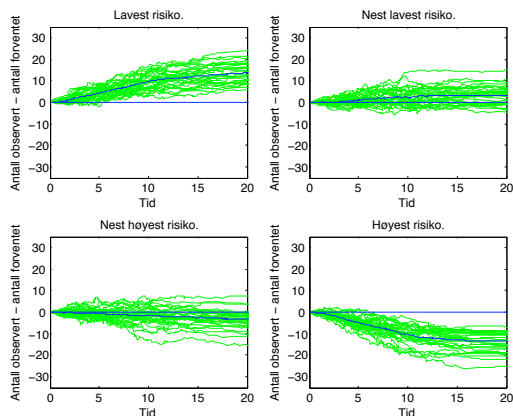
Figur 5.13 viser plot av de estimerte martingal residual prosesser når effekten av den manglende kovariaten er like stor som de andre kovariatene. For gruppen av individer som har lavest estimert risiko (øverst til venstre) ser vi at modellen estimerer for lav risiko. For gruppen av individer med høyest estimert risiko (nederst til høyre) ser vi at modellen estimerer for mange dødsfall i forhold til hva vi observerer. Differansen mellom observert og estimert antall hendelser er forholdsvis liten. Ved tid 20 er gjennomsnittlig avvik for gruppen med lavest risiko noe over 10, mens det for gruppen med høyest risiko er noe under  $-10$ . For de to resterende gruppene er gjennomsnittet av de estimerte martingal residual prosessene svært små. Figur 5.14 viser estimerte martingal residual prosesser når effekten av den manglende kovariaten er større enn effekten av de to andre ( $\beta_3 = 2$ ). Vi kjenner igjen de samme avvikene som i figur 5.14, men størrelsen på avvikene er nå noe

Tabell 5.11: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Manglende kovariat med liten effekt.

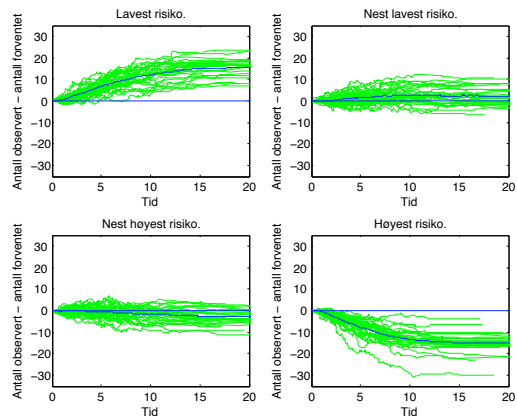
Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.900	0.016	-186.041	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.075	0.039	-1.920	0.055
4RiskR	0.023	0.038	0.618	0.537
8RiskR	0.007	0.038	0.171	0.864
4RiskKasus	0.037	0.038	0.973	0.331
8RiskKasus	0.078	0.038	2.113	0.035
2Tid2RiskR	-0.065	0.039	-1.656	0.098
80%c	-0.041	0.022	-1.864	0.062
90%c	-0.003	0.022	-0.122	0.903
KasusKontroll	-0.027	0.016	-1.761	0.078
1Kontroll	-0.060	0.016	-3.847	0.000

Tabell 5.12: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Manglende kovariat med stor effekt.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.892	0.016	-185.912	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.011	0.038	-0.283	0.777
4RiskR	-0.023	0.038	-0.591	0.555
8RiskR	-0.052	0.039	-1.349	0.177
4RiskKasus	-0.009	0.038	-0.247	0.805
8RiskKasus	0.106	0.037	2.907	0.004
2Tid2RiskR	0.034	0.038	0.908	0.364
80%c	-0.006	0.022	-0.261	0.794
90%c	-0.002	0.022	-0.113	0.910
KasusKontroll	0.019	0.016	1.245	0.213
1Kontroll	-0.041	0.016	-2.641	0.008



Figur 5.13: Estimerte martingal residualprosesser. Manglende kovariat. Alle variable har like stor effekt ( $\beta_3 = 1$ ). Simulert fra Cox-modellen, tilpasset Cox-modellen. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingal residualprosessene for 30 tilfeldige kjøring (grå) og for gjennomsnittet av 1000 kjøring (svart).



Figur 5.14: Estimerte martingal residualprosesser. Manglende kovariat. Effekten av manglende kovariat stor ( $\beta_3 = 2$ ). Simulert fra Cox-modellen, tilpasset Cox-modellen. Andelen sensurerte levetider er 80%, 3 kontroll er samlet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingal residualprosessene for 30 tilfeldige kjøring (grå) og for gjennomsnittet av 1000 kjøring (svart).

større. Ved tid 20 blir størrelsen på den estimerte martingal residual prosessen nærmere 20 for gruppen individer med lavest risiko, mens den nærmer seg  $-20$  for gruppen med høyest risiko. For de to andre gruppene passer modellen godt også her.

I en virkelig kasus-kontroll studie antar vi at fagfolk vil være med å utforme studien og sørge for at viktige variable inkluderes i studien. Det virker derfor lite sannsynlig at variable med svært stor effekt er utelatt. Derfor har vi ikke undersøkt teststyrken når den manglende kovariaten har enda større effekt.

Konklusjonen av analysene i dette avsnittet er at goodness-of-fit testen (3.27) har svært liten evne til å avsløre at en kovariat mangler fra den tilpassete modellen. Dette gjelder selv når den utelatte kovariaten har høyere effekt enn de to som er inkludert i modellen.

### 5.5.6 Feil koding av kovariat.

I avsnitt 4.3.2 argumenterte vi for at det i noen situasjoner kunne være hensiktsmessig å transformere en kovariat. I medisinske studier viser det seg for eksempel ofte fornuftig å log-transformere en kovariat. Dette er i såfall fordi effekten av kovariaten avtar etterhvert som verdien til variabelen øker.

I samme avsnitt ble det også redgjort for hvordan vi kunne utvide simuleringsalgoritmen vår til å generere simulerte levetider der effekten av en av kovariatene er log-transformert.

Tabell 5.13: Resultater fra logistisk regresjonsanalyse uten interaksjonseffekter. Feil koding av kovariat.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.917	0.016	-185.623	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.025	0.039	-0.642	0.521
2Tid4Kov	0.023	0.038	0.607	0.544
4RiskR	-0.013	0.039	-0.340	0.734
8RiskR	-0.006	0.038	-0.151	0.880
4RiskKasus	0.047	0.038	1.236	0.217
8RiskKasus	-0.025	0.039	-0.653	0.514
80%c	-0.018	0.022	-0.803	0.422
90%c	-0.026	0.022	-1.151	0.250
KasusKontroll	0.017	0.016	1.065	0.287
1Kontroll	-0.046	0.016	-2.922	0.003

I dette avsnittet vil vi underøke styrken til goodness-of-fit testen (3.27) i situasjoner der en av kovariatene er log-transformert. Vi vil simulere levetider der en av kovariatene er log-transformert og tilpasse en modell der kovariaten er lineært kodet, og undersøke hvor effektiv testen er til å avsløre at feil modell er tilpasset i denne situasjonen.

Figur 4.5 viser den log-transformerte kovariaten  $X_1'$  for ulike verdier av parameteren  $b$  fra (4.13). Jo høyere verdi av  $b$ , jo mer avviker  $X_1'$  fra  $X_1$ . Vi ser at for  $b = 30$  blir verdien til  $X_1'$  lik 0.8 når  $X_1 = 0.5$ . Dette er situasjonen vi har valgt å undersøke i dette avsnittet.

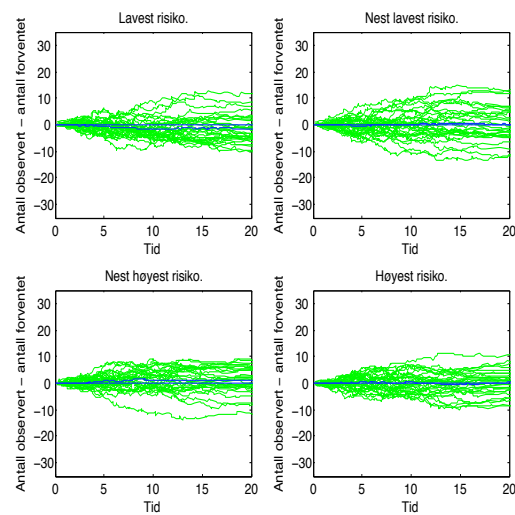
Tabell A.17 i vedlegget viser at teststyrken er svært lav i denne situasjonen. Styrken ser ut til å være lav i alle situasjoner, selv om den kanskje er noe høyere når grupperingen for testen gjøres på bakgrunn av risiko.

Resultatene av en regresjonsanalyse av simuleringsresultatene uten interaksjonseffekter er gitt i tabell 5.13. Den gjennomsnittlige teststyrken estimeres til 5.13%. Kun effekten av antall kontroller er signifikant på 1% nivå, og denne effekten er svært liten.

Regresjonsanalysen med alle 1. ordens intraksjoner er gitt i tabell A.16 i vedlegget. Her estimeres den gjennomsnittlige teststyrken til 4.9%. Kun en samspilleffekt blir signifikant på 1% nivå. Dette er samspilleffekten mellom grupperingen 8RiskKasus og antall kontroller. Størrelsen på denne effekten er imidlertid svært liten.

Figur 5.15 viser estimerte martingal residual prosesser. Plottene likner på da den sanne modellen ble tilpasset. Det er ingen systematiske avvik mellom observerte og estimerte antall hendelser.

I dette avsnittet har vi sett at goodness-of-fit testen (3.27) ikke klarer å avsløre at feil modell er tilpasset, når en kovariat er feilkodet på det nivået den er her. Plot av martingal residual prosessene viser at modellen avviker systematisk fra virkeligheten. Mistilpasningen av modellen er med andre ord for liten til å oppdages av testen.



Figur 5.15: Estimerte martingale residualprosesser. Feil koding av kovariat. Andelen sensurerte levetider er 80%, 3 kontroller er samplet per kasus ved enkel nøstet kasus kontroll. Figurene viser de estimerte martingale residualprosessene for 30 tilfeldige kjøring (grå) og for gjennomsnittet av 50 kjøring (svart).



## Kapittel 6

# Diskusjon

I dette avsluttende avsnittet vil vi oppsummere og diskutere hovedresultatene samt se videre.

I kapittel 5 undersøkte vi signifikansnivå og teststyrke til goodness-of-fit testen (3.27) i ulike situasjoner. Vi fant i avsnitt 5.4 at signifikansnivået til testen holder. Dette er et hovedresultat i studien vår. Testobservatorene viser ingen avvik fra nullfordelingen når den sanne modellen tilpasses data. Signifikansnivået til testen avhenger derfor ikke av risikofunksjon, andel sensurering, samplingdesign eller antall kontroller som samples ved hvert hendelsestidspunkt. Hvilken gruppering som benyttes for testen påvirker heller ikke signifikansnivået.

I avsnitt 5.5 undersøkte vi styrken til testen i ulike situasjoner. Først tilpasset vi modeller med feil risikofunksjon til data i avsnitt 5.5.1. Selv om eksess relativ risiko modellen er lineær mens Cox-modellen er krum, fant vi at testen hadde liten evne til å avsløre at feil modell var valgt når forskjellen på de to modellene var så liten som i avsnitt 5.5.1. Dette gjaldt også uansett risikofunksjon, andel sensurering, samplingdesign eller antall kontroller som ble samplet. Heller ikke mellom de ulike grupperingene for testen var det nevneverdige forskjeller i styrken til goodness-of-fit testen.

I avsnitt 5.5.2 gjorde vi forskjellen på modellene større. Dette førte til noe høyere styrke, men fortsatt var styrken lav. Effekten av de ulike faktorene vi varierte i simuleringene ble nå tydeligere. Andelen sensurerte levetider viste seg å ha effekt på teststyrken, som ble høyere for lavere andel sensurerte levetider. Også risikofunksjon var en viktig faktor. Når Cox-modellen ble tilpasset data simulert fra eksess relativ risiko modellen, ble styrken høyere enn i det motsatte tilfellet. Eksess relativ risiko modellen fikk heller ikke noen særlig økning i styrken med lavere andel sensur. Mellom grupperingene for testen viste det seg imidlertid fortsatt å være små forskjeller i styrke.

I de neste avsnittene simulerte vi levetider med tidsavhengig effekt av en kovariat. I avsnitt 5.5.3 så vi hvordan teststyrken var forholdsvis lav når effekten av kovariaten endret seg forholdsvis lite med tiden. Her utmerket imidlertid de grupperingene for testen som inkluderer tid seg i positiv retning. Når grupperingen for testen ble gjort etter tid, ble

styrken høyere enn når grupperingen ble gjort bare etter risiko. Også her var det en positiv effekt knyttet til lavere andel sensurering, men denne kom først og fremst testene der grupperingen var gjort etter tid til gode. Likevel ble styrken aldri høyere enn 12%. Dette var ved grupperingen 2Tid2Kov, 80% sensurering og 3 kontroller ble samplet ved hvert hendelsestidspunkt ved counter-matching.

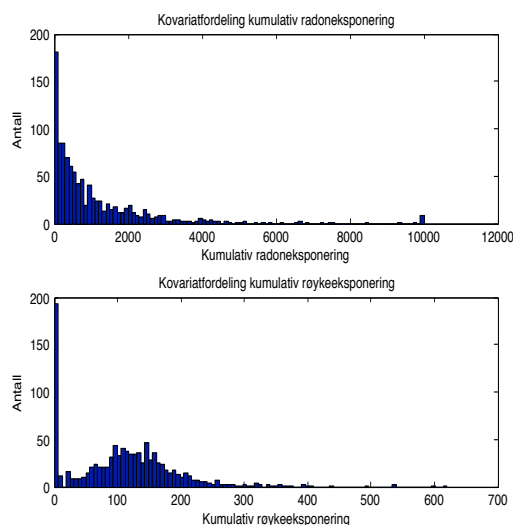
I avsnitt 5.5.4 økte vi tidsavhengigheten. Teststyrken økte uansett hvilken gruppering som ble brukt for testen, men bare i liten grad da grupperingen for testen ble gjort kun etter risiko. For disse testene ble styrken omlag like høy som i avsnitt 5.5.2 da feil risikofunksjon ble tilpasset, og det var stor forskjell på disse. Når tid ble brukt i grupperingen til testen, ble styrken nå langt høyere, særlig for lav andel sensurerte levetider. Testen viste seg dermed forholdsvis effektiv for å avsløre tidsavhengige effekter når grupperingen for testen gjøres etter tid. Dette er et av hovedresultatene i studien vår.

For testene som brukte tid i grupperingen ble også effekten av mindre andel sensurerte levetider stor. Denne effekten var mindre for testene hvor grupperingen ble gjort etter risiko. Plot av estimerte martingal residual prosesser viste seg å være effektive for å avsløre at avviket skyldes tidsavhengig effekt av en kovariat. I figur 5.12 ser man tydelig hvordan avviket mellom observert og estimert antall dødsfall øker etterhvert som effekten av kovariaten øker. De estimerte martingal residual prosessene da feil modell var tilpasset, viste ingen slike mønster. Her var endringen i de estimerte martingal residualprosessene mer lineær over tid. Dette er et nyttig resultat som kan brukes for å feildiagnostisere modeller tilpasset på virkelige levetidsdata.

I avsnittene 5.5.5 simulerte vi levetider med tre kovariater, mens vi tilpasset en modell med bare to kovariater. Her så vi at goodness-of-fit testen ikke klarte å avsløre at modellen som ble tilpasset, manglet en kovariat. Dette gjaldt også da vi lot den manglende kovariaten ha større effekt enn de to kovariatene som ble inkludert i modellen. I avsnitt 5.5.6 log-transformerte vi en av de to kovariatene i simuleringene og tilpasset en modell hvor begge kovariatene var lineært kodet. Her så vi at testen ikke klarte å avsløre at feil modell var valgt når en av kovariatene var feilkodet. Disse resultatene gjelder uavhengig av andel sensurering, antall kontroller eller samplingdesign. Hvilken gruppering som brukes for testen hadde heller ikke betydning for teststyrken i disse situasjonene.

Mange andre interessante situasjoner kunne ha vært studert, men det var det ikke tid til innen rammen av denne oppgaven. Det hadde vært interessant å undersøke hvor store avvikene må være mellom modellen brukt i simuleringene og den som er tilpasset, før testen vil avsløre med rimelig grad av sikkerhet at feil modell er valgt. Man kan øke feilkodingen i avsnitt 5.5.6 og undersøke hvor stor denne må være før testen oppdager at feil modell er valgt. Kanskje er det ikke nok å øke parameteren  $b$  i (4.13). Andre former for feilkoding kan i såfall konstrueres. Det ville også være interessant å undersøke hvor stor effekten av den manglende kovariaten i avsnitt 5.5.5 må være før teststyrken blir høyere, eller hvor stor forskjell det må være på risikofunksjonene før testen skjønner at feil modell er valgt.

Det ville være spennende å studere om andre grupperinger for testen enn de vi har tatt med i denne oppgaven fikk høyere teststyrke i noen situasjoner. Spesielt ville det være interessant å legge større vekt på tid i grupperingen. Man kunne delt inn i flere enn to tidgrupper som vi har brukt i denne oppgaven. Siden det har vist seg viktig å ta tid



Figur 6.1: Fordelingen til kovariatene fra eksempel 5. Kumulativ radoneksponering øverst og kumulativ røykeeksponering nederst.

med i grupperingen for testen for å avsløre tidsavhengig effekt av kovariat, ville det vært spennende å undersøke om denne grupperingen ga enda høyere styrke i situasjonen med tidsavhengig effekt av kovariat eller andre situasjoner.

For urangruvearbeiderne i eksempel 5 ble det mye større forskjell på Cox og eksess relativ risiko modellene. Større avvik enn det vi hadde i avsnittene 5.5.1 og 5.5.2 forekommer i altså i praksis. Figur 6.1 viser fordelingen til kovariatene til urangruvearbeiderne. Vi ser at fordelingene er svært ulike de vi har brukt i våre simuleringer. Fordelingene er svært skjeve. Det er vanskelig å vite hva som er vanlige kovariatfordelinger i kasus-kontroll studier. For å finne ut det må man undersøke flere virkelige datasett. Kanskje har kovariatfordelingene vi har brukt i simuleringene for lite variasjon i forhold til det man observerer i virkeligheten. I såfall kan det tenkes at vi har stilt kji-kvadrat testen på prøver som er vanskeligere enn de den vil bli stilt ovenfor i virkeligheten. I virkelige studier vil man antagelig også støte på kombinasjoner av flere, eller kanskje til og med alle, situasjonene vi har tatt for oss i denne oppgaven.

Kji-kvadrat testen (3.27) er en “omnibus test” som tester om riktig modell er tilpasset data uten noen antakelse om hva som eventuelt er galt med modellen. Alternativt kan mer spesifikke tester som sjekker mot bestemte avvik fra modellen benyttes. For å teste om tidsavhengig effekt av en kovariat er tilstede, kan man ta med en kovariat med tidsavhengig effekt. For eksempel kan man ta med en tredje kovariat  $X_3(t) = X_1 \cdot \log t$ . Deretter tester man om koeffisienten for denne,  $\beta_3$  er signifikant forskjellig fra null. Å teste om en kovariat er riktig kodet kan gjøres på omtrent samme måte. Ved for eksempel å ta med den transformerte kovariaten  $X_3 = X_1^2$  og så teste om  $X_3$  har signifikant effekt, kan man undersøke om kovariaten har denne feilkodingen i modellen.

Med mer spesifikke tester kan man sjekke modellen for bestemte avvik, som tidsavhengig

effekt av kovariat eller feil koding av kovariat. Slike tester har større styrke mot bestemte avvik fra modellen enn vår  $\chi^2$ -test. Ulempen med slike tester er at man må teste for en rekke ulike avvik. Modellen kan være gal uten at testen avslører dette, fordi man ikke tester mot riktig modellavvik. Her har  $\chi^2$  goodness-of-fit testen (3.27) en fordel. Denne testen vil kunne oppdage modellavvik uansett hva disse skyldes. Ulempen med slike tester er at man, hvis testen gir forkastning, ikke får innsikt i hva som eventuelt er feil med modellen, bare at den er feil. Dessuten har altså slike “omnibus tester” lavere teststyrke enn spesifikke tester. Med en mer spesifikk test får man også innsikt i hva som var galt med modellen.

Den lave estimerte teststyrken i de fleste situasjoner tyder på at  $\chi^2$ -testen har dårlig styrke for å avsløre avvik som er av interesse. For kasus-kontroll studier kan det derfor hende at en ikke kan bruke en slik “omnibus test” som  $\chi^2$ -testen, men i stedet må bruke tester som har større styrke mot bestemte avvik fra modellen.

## Tillegg A

## Vedlegg

I vedlegget finnes lister som viser andel tester som gir forkastning for hver av situasjonene som er omtalt i kapittel 5. Enkelte av disse er delt i to fordi de ellers ikke ville få plass på en side. Dette er i såfall forklart i tabellen.

Rekkefølgen på tabeller og resultater fra regresjonsanalysene er den samme som rekkefølgen på analysene av simuleringsresultatene i kapittel 5.

Tabell A.1: Simuleringsresultater når modellen stemmer. Testens signifikansnivå. Andel tester som gir forkastning. Tabellen fortsetter i tabell A.2

2Tid2Kov								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	5.3	4.9	4.8	4.5	5.6	4	4.7	4.5
0.90	4.8	4.7	5.6	5	5.2	6	6.1	4.8
0.95	4.4	4.5	5.4	5.4	5.6	4	7	7
4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	4.8	7.8	6.5	4.9	4.4	5.3	4.7	4.9
0.90	5.3	5.1	4.5	4.6	4.6	5.7	4.5	5.3
0.95	5	4.4	5.3	6	6	4.7	5.6	5.6
4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	4.8	7.8	6.5	4.9	4.4	5.3	4.7	4.9
0.90	5.3	5.1	4.5	4.6	4.6	5.7	4.5	5.3
0.95	5	4.4	5.3	6	6	4.7	5.6	5.6
8RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	4	6.1	5.4	4.5	4.5	4.6	4.7	5.2
0.90	5	5.6	4.2	4.6	5.5	6.9	4.9	5.4
0.95	4.7	4.9	4.1	5.9	3.9	5.8	4	4
4RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	4.8	5.6	6	4.8	6	5.1	5.8	5.3
0.90	5.4	5.4	4.1	5.7	5.9	5.5	5.7	4.4
0.95	5.6	6.6	4.5	5.7	4.8	7.2	5.6	5.6
8RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	5	6.2	5.4	6.1	6.2	5.8	4.5	6.3
0.90	5.3	7.1	4.6	7.2	6.1	7.4	5.6	7.4
0.95	4.8	9.3	3.7	8.3	3.4	9.8	5.1	5.1

Tabell A.2: Fortsettelse av tabell A.1 Simuleringsresultater når modellen stemmer. Andel tester som gir forkastning. Testens signifikansnivå.

2Tid2RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	5	6	5	4.9	4.6	4.5	4.4	5.5
0.90	4.9	5.4	5.5	5.2	5.4	7.2	4.6	3.9
0.95	6	4.5	5.5	6.2	6.2	5.3	4.3	4.3
2Tid4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.80	3.4	6.5	5.5	5.1	4.7	5.3	4.1	4.1
0.90	4.8	5.1	5.4	4.8	4.2	5.9	4.4	4.4
0.95	4.3	4.9	4.1	5.2	4.5	4.3	3.6	3.6

Tabell A.3: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Feil modell tilpasset. Liten forskjell på modellene.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Intercept	-2.918	0.011	-267.838	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.032	0.027	-1.219	0.222
4RiskR	0.038	0.026	1.432	0.152
8RiskR	-0.019	0.027	-0.705	0.481
4RiskKasus	-0.019	0.027	-0.705	0.481
8RiskKasus	0.106	0.026	4.113	$3.90 \cdot 10^{-5}$
2Tid2RiskR	0.003	0.027	0.117	0.907
Cox	0.023	0.011	2.159	0.031
80%c	-0.014	0.015	-0.916	0.360
90%c	0.018	0.015	1.161	0.246
KasusKontroll	0.000	0.011	0.034	0.973
1Kontroll	-0.067	0.011	-6.142	$8.14 \cdot 10^{-10}$
2Tid2Kov:1Kontroll	0.107	0.027	3.978	$6.95 \cdot 10^{-5}$
4RiskR:1Kontroll	0.022	0.026	0.849	0.396
8RiskR:1Kontroll	-0.014	0.027	-0.531	0.595
4RiskKasus:1Kontroll	-0.014	0.027	-0.531	0.595
8RiskKasus:1Kontroll	-0.147	0.026	-5.681	$1.34 \cdot 10^{-8}$
2Tid4RiskR:1Kontroll	0.049	0.027	1.834	0.067

Tabell A.4: Andel tester som gir forkastning. Feil linkfunksjon tilpasset. Liten forskjell på modellene. Eksess relativ risiko betyr simulert fra eksess relativ risiko modellen, tilpasset Cox-modellen, og motsatt. Tabellen fortsetter i tabell A.5

2Tid2Kov								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.1	4	4.5	4.5	5.7	4.5	4.4	5
0.9	4.8	3.9	6.1	6.3	5.3	6.5	6.4	4.3
0.95	5.6	5.2	5.9	5.8	5.8	4.4	5.7	5.7
4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	4.5	7.2	6.9	5.3	5.3	6.3	5.4	5.8
0.9	5.7	6.2	4.6	5.2	5.1	6.4	5.8	5.5
0.95	5.2	4.2	5.3	6.2	5	5.8	5.7	5.7
8RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.1	5.5	5	5.9	4.9	5.8	4.9	5.5
0.9	5.6	5.9	3.9	4.9	5.2	7.2	6	6
0.95	4.8	4.6	4.6	5.3	4.1	6.2	4.6	4.6
4RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5	6.4	5	4.7	6.1	5.5	5.8	5.3
0.9	5.7	5	3.8	5	6	5.5	5.6	4.9
0.95	5.6	6.2	4.8	6.7	4.4	6.3	5	5
8RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5	6.2	5.4	6	6.1	6.8	4.5	5.7
0.9	5.1	6.9	4.4	6.9	5.8	7.4	5	7.1
0.95	5.2	9.2	3.8	9.1	3	8.8	4.8	4.8



Tabell A.5: Fortsettelse av tabell A.5 Andel tester som gir forkastning. Feil linkfunksjon tilpasset. Liten forskjell på modellene. Eksess relativ risiko betyr simulert fra eksess relativ risiko modellen, tilpasset Cox-modellen, og motsatt.

2Tid2RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	4.7	5.7	5.7	4.5	5	6.4	5.2	6.3
0.9	4.9	5	5.8	5.2	5.7	6.6	5.4	4.6
0.95	5.2	4	5.4	6	5.5	6	5.5	5.5
2Tid4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	4.1	5.6	6.4	5.7	4.4	6	3.8	5.2
0.9	5.2	5.6	5.1	5.3	3.7	6.4	4.9	4.9
0.95	4.1	4.4	4.7	4.9	5	4.4	4.8	4.8

Tabell A.6: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Feil modell tilpasset. Stor forskjell på modellene.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.737	0.010	-263.629	$2 \cdot 10^{-16}$
2Tid2Kov	0.053	0.025	2.134	0.033
4RiskR	0.031	0.025	1.226	0.220
8RiskR	0.044	0.025	1.766	0.077
4RiskKasus	0.021	0.025	0.834	0.404
8RiskKasus	0.053	0.025	2.108	0.035
2Tid2RiskR	-0.101	0.028	-3.929	$8.52 \cdot 10^{-5}$
Cox	0.188	0.010	18.183	$2 \cdot 10^{-16}$
80%c	0.148	0.014	10.561	$2 \cdot 10^{-16}$
90%c	-0.016	0.014	-1.085	0.278
sampleDesign	-0.025	0.010	-2.508	0.012
1Kontroll	-0.080	0.010	-7.863	$3.75 \cdot 10^{-15}$
2Tid2Kov:met1	0.119	0.025	4.828	$1.38 \cdot 10^{-6}$
4RiskR:met1	0.066	0.025	2.635	0.008
8RiskR:met1	0.047	0.025	1.893	0.058
4RiskKasus:met1	-0.020	0.025	-0.807	0.420
8RiskKasus:met1	-0.106	0.025	-4.375	$1.21 \cdot 10^{-5}$
2Tid2RiskR:met1	-0.083	0.026	-3.173	0.002
2Tid2Kov :andelC1	0.020	0.032	0.610	0.542
4RiskR:andelC1	0.012	0.033	0.365	0.715
8RiskR:andelC1	0.121	0.033	3.724	0.000
4RiskKasus:andelC1	0.012	0.033	0.370	0.711
8RiskKasus:andelC1	0.054	0.033	1.661	0.097
2Tid2RiskR:andelC1	-0.159	0.036	-4.429	$9.46 \cdot 10^{-6}$
2Tid2Kov :andelC2	0.031	0.034	0.910	0.363
4RiskR:andelC2	-0.031	0.035	-0.874	0.382
8RiskR:andelC2	-0.003	0.035	-0.095	0.924
4RiskKasus:andelC2	0.010	0.035	0.281	0.779
8RiskKasus:andelC2	-0.021	0.035	-0.622	0.534
2Tid2RiskR:andelC2	0.016	0.036	0.453	0.651
2Tid2Kov :nKontr1	0.041	0.024	1.717	0.086
4RiskR:nKontr1	0.044	0.024	1.826	0.068
8RiskR:nKontr1	-0.010	0.024	-0.431	0.667
4RiskKasus:nKontr1	-0.043	0.024	-1.770	0.077
8RiskKasus:nKontr1	-0.154	0.024	-6.309	$2.80 \cdot 10^{-10}$
2Tid2RiskR:nKontr1	0.056	0.026	2.165	0.030
met1:andelC1	0.147	0.014	10.533	$2 \cdot 10^{-16}$
met1:andelC2	-0.022	0.014	-1.514	0.130
met1:sample1	-0.034	0.010	-3.352	0.001
met1:nKontr1	-0.035	0.010	-3.442	0.001

Tabell A.7: Andel tester som gir forkastning. Feil linkfunksjon tilpasset. Stor forskjell på modellene. Eksess relativ risiko betyr simulert fra eksess relativ risiko modellen, tilpasset Cox-modellen, og motsatt. Tabellen fortsetter i tabell A.8

2Tid2Kov								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	4.5	5.3	4.8	4.7	13	16	10	13
0.9	6.2	5.1	4.6	5.4	9.6	11	8.1	8.4
0.95	4.5	5.7	5.1	4.4	8.6	6.5	4.3	4.3
4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.7	5.5	5.7	4.3	12	13	10	11
0.9	4.7	5.8	5.4	4.4	7	9.9	6	9.4
0.95	3.5	5.8	5.6	5.7	7.4	6.7	7.1	7.1
8RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	6	6.6	5.3	6.7	14	15	8.8	14
0.9	4.5	6.3	6.7	4.8	6.8	9.4	6.1	10
0.95	4.1	5.4	4.7	3.6	5.9	7	6	6
4RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	4.4	6.8	5.7	5.6	11	12	7.8	12
0.9	5.1	6.4	5.2	5.6	6.5	9.2	7	8.7
0.95	4.5	6.1	5.5	6.5	4.9	7.3	4.6	4.6
8RiskKasus								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.2	6.9	5.3	7.2	10	14	8.9	13
0.9	5.6	7.8	5	7.6	5.9	9.8	4.1	8.4
0.95	4	8.1	4.8	9.6	2.9	6.7	3.7	3.7

Tabell A.8: Fortsettelse av tabell A.7 Andel tester som gir forkastning. Feil linkfunksjon tilpasset. Stor forskjell på modellene. Eksess relativ risiko betyr simulert fra eksess relativ risiko modellen, tilpasset Cox-modellen, og motsatt.

2Tid2RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.2	4.2	7	4.3	6	7.5	5.8	7.7
0.9	3.8	5.4	4.9	6.2	5.9	7.1	6.2	7.7
0.95	5.4	4.9	6.3	5.9	6.3	7	6.6	6.6
2Tid4RiskR								
Risikofunksjon Samplingdesign Antall kontroller	Eksess relativ risiko				COX			
	CM		KK		CM		KK	
	1	3	1	3	1	3	1	3
0.8	5.8	4.4	5.1	3.8	9	9.5	7.7	8.4
0.9	4.7	5.5	5.3	4.3	4.9	8.6	5.8	7.5
0.95	4.1	5.5	6.1	5	6.7	5.6	5.6	5.6

Tabell A.9: Andel tester som gir forkastning. Tidsavhengig effekt av kovariat. Positiv, liten effekt.

2Tid2Kov					4RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	12	9	8.2	9.8	0.8	5.9	4.4	4.2	6
0.9	6.4	7.6	6.2	5.5	0.9	5.4	3.9	5.3	6.5
0.95	5.7	6.9	5.3	5.3	0.95	5.5	5.3	5.1	5.1
8RiskR					4RiskKasus				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	6.3	4.7	4.7	5.8	0.8	6.3	5.3	4.9	5.3
0.9	5.3	5.2	3.9	6.3	0.9	5.2	4.5	4.9	5
0.95	5.1	4.2	4.5	4.5	0.95	5.3	5.9	6.7	6.7
8RiskKasus					2Tid2RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	7.2	5.6	5.3	6.5	0.8	9.8	8.7	6	8.4
0.9	6	5.9	4.6	7.1	0.9	7.3	6.7	5.7	6.5
0.95	4.8	8.8	5.5	5.5	0.95	4.9	5.4	4.4	4.4
2Tid4RiskR									
Risikofunksjon	Cox								
Samplingdesign	CM		KK						
Antall kontroller	1	3	1	3					
0.80	9	6.6	5.5	7.4					
0.90	6.6	4.5	5.3	5.1					
0.95	4.9	5.3	4.5	4.5					

Tabell A.10: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Tidsavhengige effekt av kovariat. Stor positiv effekt.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-1.975	0.011	-173.251	$< 2 \cdot 10^{-16}$
2Tid2Kov	1.137	0.018	61.504	$< 2 \cdot 10^{-16}$
4RiskR	-0.718	0.033	-21.774	$< 2 \cdot 10^{-16}$
8RiskR	-0.672	0.032	-20.770	$< 2 \cdot 10^{-16}$
4RiskKasus	-0.768	0.033	-22.966	$< 2 \cdot 10^{-16}$
8RiskKasus	-0.668	0.032	-20.809	$< 2 \cdot 10^{-16}$
2Tid2RiskR	0.935	0.019	48.335	$< 2 \cdot 10^{-16}$
80%c	0.285	0.015	18.679	$< 2 \cdot 10^{-16}$
90%c	-0.031	0.016	-1.914	0.056
KasusKontroll	-0.202	0.011	-17.889	$< 2 \cdot 10^{-16}$
1Kontroll	-0.062	0.011	-5.515	$3.49 \cdot 10^{-8}$
2Tid2Kov:80%c	0.168	0.024	6.944	$3.80 \cdot 10^{-12}$
4RiskR :80%c	-0.071	0.044	-1.614	0.107
8RiskR :80%c	-0.126	0.043	-2.909	0.004
4RiskKasus:80%c	-0.160	0.046	-3.515	0.000
8RiskKasus :80%c	-0.279	0.045	-6.252	$4.05 \cdot 10^{-10}$
2Tid2RiskR:80%c	0.197	0.025	7.847	$4.26 \cdot 10^{-15}$
2Tid2Kov:90%c	0.109	0.025	4.284	$1.84 \cdot 10^{-5}$
4RiskR :90%c	0.008	0.046	0.170	0.865
8RiskR :90%c	-0.052	0.046	-1.128	0.259
4RiskKasus:90%c	0.022	0.047	0.472	0.637
8RiskKasus :90%c	-0.031	0.046	-0.681	0.496
2Tid2RiskR:90%c	-0.006	0.027	-0.222	0.824
2Tid2Kov:KasusKontroll	-0.048	0.018	-2.663	0.008
4RiskR :KasusKontroll	0.044	0.033	1.356	0.175
8RiskR :KasusKontroll	-0.013	0.032	-0.393	0.695
4RiskKasus:KasusKontroll	0.087	0.033	2.614	0.009
8RiskKasus :KasusKontroll	0.071	0.032	2.198	0.028
2Tid2RiskR:KasusKontroll	-0.064	0.019	-3.369	0.001
2Tid2Kov:1Kontroll	-0.067	0.018	-3.746	0.000
4RiskR :1Kontroll	0.094	0.032	2.903	0.004
8RiskR :1Kontroll	0.074	0.032	2.329	0.020
4RiskKasus:1Kontroll	0.050	0.033	1.496	0.135
8RiskKasus :1Kontroll	0.012	0.032	0.387	0.699
2Tid2RiskR:1Kontroll	-0.081	0.019	-4.325	$1.52 \cdot 10^{-5}$
KasusKontroll:1Kontroll	-0.124	0.009	-13.854	$< 2 \cdot 10^{-16}$

Tabell A.11: Andel tester som gir forkastning. Tidsavhengig effekt av kovariat. Positiv stor effekt.

2Tid2Kov					4RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	83	89	44	67	0.8	13	7.8	6.3	7.6
0.9	62	60	26	48	0.9	7.8	8	5.3	5.9
0.95	33	33	16	16	0.95	7.4	5.2	4.8	4.8

8RiskR					4RiskKasus				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	13	8.1	5.7	7.3	0.8	9.4	8.5	5.1	6.6
0.9	9.1	7.1	4.8	5.9	0.9	7.8	6.4	5.1	6.6
0.95	8.9	7.1	4.1	4.1	0.95	6.2	5.5	5.8	5.8

8RiskKasus					2Tid2RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	9.5	7.5	5.9	6.2	0.8	75	76	32	58
0.9	7.9	7.5	5.1	6.7	0.9	43	45	21	34
0.95	8.1	8.3	5.4	5.4	0.95	29	30	14	14

2Tid4RiskR				
Risikofunksjon	COX			
Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3
0.8	71	68	28	50
0.9	35	36	15	28
0.95	23	23	11	11

Tabell A.12: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Mangler kovariat. Alle kovariater lik effekt. ( $\beta_3 = 1$ )

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.957	0.016	-189.479	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.068	0.039	-1.747	0.081
4RiskR	0.023	0.038	0.602	0.547
8RiskR	0.009	0.038	0.233	0.815
4RiskKasus	0.036	0.038	0.964	0.335
8RiskKasus	0.061	0.038	1.617	0.106
2Tid2RiskR	-0.059	0.039	-1.501	0.133
80%c	-0.042	0.022	-1.882	0.060
90%c	$7.525 \cdot 10^{-5}$	0.022	0.003	0.997
KasusKontroll	-0.026	0.016	-1.665	0.096
1Kontroll	-0.055	0.016	-3.557	0.000
2Tid2Kov:1Kontroll	0.041	0.039	1.054	0.292
4RiskR:1Kontroll	-0.017	0.038	-0.453	0.651
8RiskR:1Kontroll	0.017	0.038	0.435	0.663
4RiskKasus:1Kontroll	$-5.144 \cdot 10^{-4}$	0.038	-0.014	0.989
8RiskKasus:1Kontroll	-0.126	0.038	-3.367	0.001
2Tid2RiskR:1Kontroll	0.023	0.039	0.584	0.559
80%c:1Kontroll	-0.044	0.022	-1.999	0.046
90%c:1Kontroll	0.058	0.022	2.622	0.009



Tabell A.13: Andel tester som gir forkastning. Mangler kovariat med liten effekt.

2Tid2Kov					4RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.4	4.8	3.4	4.5	0.8	4.4	6.6	4.8	5.6
0.9	6.3	4.9	5.1	4.9	0.9	5.3	5.3	4	4.6
0.95	4.7	5.3	4.9	4.9	0.95	4.7	6.3	6.5	6.5
8RiskR					4RiskKasus				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.7	5.9	5.3	4.8	0.8	4.6	6	4.9	5
0.9	5.6	5.7	4.6	5.5	0.9	7.1	5	4	5.4
0.95	4.4	4.9	5.7	5.7	0.95	4.9	5.7	5.1	5.1
8RiskKasus					2Tid2RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.8	7.4	3.5	5.6	0.8	4.4	5.5	4.6	5.1
0.9	5.2	5.9	4.5	6.6	0.9	5.5	4.9	4.3	4.9
0.95	4.3	7.8	5.3	5.3	0.95	4.3	5.1	5.3	5.3
2Tid4RiskR									
Risikofunksjon	COX								
Samplingdesign	CM		KK						
Antall kontroller	1	3	1	3					
0.8	5.3	5.8	4	4.8					
0.9	6.4	4.7	4.9	4.5					
0.95	4.9	5.2	5.9	5.9					

Tabell A.14: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Mangler kovariat. Manglende kovariat stor effekt. ( $\beta_3 = 2$ )

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.948	0.016	-189.486	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.008	0.038	-0.204	0.838
4RiskR	-0.019	0.038	-0.499	0.618
8RiskR	-0.048	0.039	-1.230	0.219
4RiskKasus	-0.009	0.038	-0.241	0.810
8RiskKasus	0.090	0.037	2.443	0.015
2Tid2RiskR	0.034	0.038	0.918	0.359
80%c	-0.005	0.022	-0.233	0.816
90%c	-0.003	0.022	-0.116	0.907
KasusKontroll	0.018	0.016	1.181	0.238
1Kontroll	-0.037	0.016	-2.410	0.016
2Tid2Kov:1Kontroll	0.051	0.038	1.352	0.176
4RiskR:1Kontroll	0.024	0.038	0.636	0.525
8RiskR:1Kontroll	0.005	0.039	0.132	0.895
4RiskKasus:1Kontroll	-0.036	0.038	-0.954	0.340
8RiskKasus:1Kontroll	-0.120	0.037	-3.241	0.001
2Tid2RiskR:1Kontroll	0.068	0.038	1.819	0.069

Tabell A.15: Andel tester som gir forkastning. Mangler kovariat med stor effekt.

2Tid2Kov					4RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	5.6	5.7	4.5	5.2	0.8	3.9	5.4	6	4.9
0.9	5.2	4.4	4.8	5.1	0.9	4.7	4.9	5.8	4.8
0.95	5.8	5.2	5.8	5.8	0.95	4.8	5.9	5.3	5.3
8RiskR					4RiskKasus				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.9	5.1	5.6	4.8	0.8	4.4	4.1	4.9	6.3
0.9	4.3	5.1	5.5	5.8	0.9	4.9	6.6	4.8	5.3
0.95	4.4	5.4	4.3	4.3	0.95	5.1	6.1	4.8	4.8
8RiskKasus					2Tid2RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.7	4.4	5.4	6.1	0.8	4.6	5.7	6.7	6.2
0.9	5.2	6.8	4.4	7.3	0.8	6	5.2	5.7	5.3
0.95	4.7	8.2	5	5	0.95	5	5	5.5	5.5
2Tid4RiskR									
Risikofunksjon	COX								
Samplingdesign	CM		KK						
Antall kontroller	1	3	1	3					
0.8	4.4	5.4	6.8	4.9					
0.9	4.1	5.2	4.2	5.5					
0.95	4.8	5.6	4.9	4.9					

Tabell A.16: Resultater fra logistisk regresjonsanalyse med signifikante interaksjonseffekter. Feil koding av kovariat.

Koefisienter	Estimat	St. Error	z-verdi	p-verdi
Konstantledd	-2.972	0.016	-189.019	$< 2 \cdot 10^{-16}$
2Tid2Kov	-0.022	0.039	-0.572	0.568
4RiskR	0.025	0.038	0.653	0.514
8RiskR	-0.010	0.039	-0.262	0.793
4RiskKasus	-0.005	0.038	-0.128	0.898
8RiskKasus	0.032	0.038	0.824	0.410
2Tid2RiskR	-0.021	0.039	-0.554	0.580
80%c	-0.017	0.022	-0.753	0.451
90%c	-0.025	0.022	-1.099	0.272
KasusKontroll	0.016	0.016	1.015	0.310
1Kontroll	-0.043	0.016	-2.726	0.006
2Tid2Kov:1Kontroll	-0.010	0.039	-0.263	0.792
4RiskR:1Kontroll	0.037	0.038	0.960	0.337
8RiskR:1Kontroll	0.008	0.039	0.196	0.845
4RiskKasus:1Kontroll	-0.025	0.038	-0.644	0.520
8RiskKasus:1Kontroll	-0.132	0.038	-3.435	0.001
2Tid2RiskR:1Kontroll	0.073	0.039	1.879	0.060

Tabell A.17: Andel tester som gir forkastning. Feil koding av kovariat.

2Tid2Kov					4RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.3	4.7	4.7	5.9	0.8	5.5	5.6	5	4.3
0.9	5	5.1	3.7	6.1	0.9	5.2	4.4	5.4	5.1
0.95	5.4	5.3	5.4	5.4	0.95	5.3	5.8	4.9	4.9
8RiskR					4RiskKasus				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	5	5.5	5.1	5.4	0.8	5.1	5.2	5.4	5
0.9	4.9	4.9	4.9	5.3	0.9	4.7	4.6	4.9	5.7
0.95	4.5	5.2	5	5	0.95	4.3	6.5	4.1	4.1
8RiskKasus					2Tid2RiskR				
Risikofunksjon	COX				Risikofunksjon	COX			
Samplingdesign	CM		KK		Samplingdesign	CM		KK	
Antall kontroller	1	3	1	3	Antall kontroller	1	3	1	3
0.8	4.8	5.9	5.7	5.7	0.8	4.2	4.6	5.1	4.8
0.9	3.5	5.9	3.8	6.9	0.9	5.3	3.7	6	4.4
0.95	4.3	6.3	4.4	4.4	0.95	5.3	5.9	5.1	5.1
2Tid4RiskR									
Risikofunksjon	COX								
Samplingdesign	CM		KK						
Antall kontroller	1	3	1	3					
0.8	5.3	4.5	4.7	3.9					
0.9	5.2	4.7	4.7	6.6					
0.95	5.7	5.3	5.4	5.4					



# Kilder

1. AALEN, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist. Med.* **12** 1569-1588.
2. AALEN, O. O., BORGAN, Ø, & GJESSING, H. K. (2008). *Event History Analysis A Process Point of View*. Springer-Verlag, New York
3. BARLOW, W. E. & PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75**, 63-74.
4. BORGAN, Ø., GOLDSTEIN, L. & LANGHOLZ, B. (1995). Methods for the Analysis of Sampled Cohort Data in the Cox Proportional Hazards Model. *Ann. Statist.* **23**, 1749-1778.
5. BORGAN, Ø. & LANGHOLZ, B. (2007). Using martingale residuals to assess goodness-of-fit for sampled risk set data. In V. NAIR (Ed.), *Advances in Statistical Modeling and Inference. Essays in Honour of Kjell A Doksum*, pp. 65-90. World Scientific Publishing, New York
6. GRØNNESBY, J. K. & BORGAN, Ø. (1996). A method for checking regression models in survival analysis based on risk score. *Lifetime Data Anal.* **2**, 315-328.
7. HOSMER, D. W. & LEMESHOW, S. (1989). *Applied Logistic Regression* John Wiley & Sons.
8. KLEIN, J. P. & MOESCHBERGER, M. L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data. 2nd ed.* Springer-Verlag, New York.
9. LANGHOLZ, B & BORGAN, Ø. (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* **82**, 69-79.
10. RICE, J. A. (1995). *Mathematical Statistics and Data Analysis. 2nd ed.* Duxbury Press, Belmont, California.